Mathematisch-Naturwissenschaftlichen Fakultät II der Humboldt Universität zu Berlin

DISSERTATION

The Influence of Task Instructions on Action Coding: Response Instruction and Response Coding

Zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Psychologie
eingereicht an der Mathematisch-Naturwissenschaftlichen Fakultät II der Humboldt Universität zu Berlin

Dipl. Psych. Dorit Wenke, geboren am 22.3.1968 in Oldenburg

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II: Prof. Dr. Elmar Kulke

Gutachter: 1. Prof. Dr. Peter A. Frensch
2. Prof. Dr. Bernhard Hommel
3. PD Dr. Iring Koch

Datum der Promotion: 12.12.2003

Acknowledgements

I want to express my gratitude towards the persons and institutions who supported the genesis of this thesis.

Special thanks go to my supervisor Peter Frensch for all the helpful discussions, suggestions, revisions, and pep talks, the support regarding all kinds of strategic and organizational matters, as well as his patience and trust during the last years.

I also want to thank my colleagues and friends from the department for helpful data discussions, experimental ideas, and motivational support when the going got tough. In this regard, I am particularly grateful to Dieter Nattkemper, Birgit Stürmer, and Michael Ziessler who shaped my thinking about coding issues, and who provided helpful comments on earlier versions of this thesis.

I also want to express my gratitude towards Ina-Kristin Saure and Annekathrin Schacht for their help in programming the experiments and in collecting and coding of data.

Thanks to Ina-Kristin Saure, Michaela Schneider, Rico Fischer, and Klaus Fenchel for proof-reading (parts of) this thesis under high time pressure.

Furthermore I would like to express my gratitude towards the members of this committee, especially Bernhard Hommel, Iring Koch, and Werner Sommer for their willingness to invest their time.

Finally, I want to mention that part of this thesis was sponsored by the "Berliner Nachwuchsförderung" (NaFöG) of the state of Berlin, without which this thesis would not have been possible.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANOVA	Analysis of Variance
CTC	Cross-Task Compatibility
DCC	Display-Control-Arrangement Correspondence
DO	Dimensional Overlap
ERTS©	Experimental Runtime System
H&M task	Hedge and Marsh task
IRI	Inter-Response Interval
LTM	Long-term Memory
MANOVA	Multivariate Analysis of Variance
ms	Milliseconds
MSe	Mean Square Error
PFC	Prefrontal Cortex
PI	% Invalid Responses
PMC	Perceptual-Motor Cortex
PRP	Psychological Refractory Period
R; R(i)	Response; Response (i)
RT	Reaction Time
S; S(i)	Stimulus; Stimulus (i)
S-R; R-R; S-S	Stimulus-Response; Response-Response; Stimulus-Stimulus
SOA	Stimulus-Onset Asynchrony
SRC	Stimulus-Response Compatibility
STM	Short-term Memory
TEC	Theory of Event Coding

Abstract

According to Monsell (1996), one of the ‘unsolved mysteries of mind’ is how exactly verbal task instructions are translated into, and are used to control behavior. The present dissertation attempts to shed some light on one aspect of this mystery, namely on how the wording of task instructions affects the codes and processes commonly associated with response selection, a processing ‘stage’ assumed to be central in action control. The main question is whether or not the response labels used in the instructions of manual two-choice responses affect how responses are coded and accessed. If instruction determines response coding, then it should be possible to demonstrate that identical tasks are performed differently if response instructions differ.

In five experiments, I manipulated response instructions for spatially organized keypress responses. Specifically, I instructed left and right keypresses on a manual task either as left vs. right or as blue vs. green keypresses and tested whether such variations in response instructions affect two different types of compatibility effects.

The first set of experiments (Experiments 1-3) used a dual task procedure that, in addition to the manual task, required either “left” vs. “right” or “blue” vs. “green” verbalizations on a concurrently performed verbal task. When responses on both the manual and the verbal task were instructed in terms of location (Experiment 1) or color (Experiment 2), then compatible responses on the two tasks (e.g., “blue” verbalizations followed by a blue keypress) were faster than incompatible responses. However, when the verbal task required “left” vs. “right” responses whereas manual keypresses were instructed as blue vs. green (Experiment 3), then no compatibility effects were observed.

The second set of experiments (Experiments 4 and 5) extended these findings by employing the same response-instruction logic to a Simon-like task, in which left and right keypress responses were arbitrarily mapped to centrally presented stimuli (letter identity). Go/No-go signals that varied in location indicated whether the prepared response was to be executed or not. Color instructions of the response keys (Experiment 5) significantly reduced the Simon effect (i.e., faster responses when response location and irrelevant Go/No-go location correspond) observed under spatial response instructions (Experiment 4).

Taken together, these results suggest that response labels used in the instruction directly determine the codes that are used to control responding, and that non-spatial coding can override spatial coding under non-spatial response instructions. The findings are discussed with

respect to their relevance for contemporary coding accounts of compatibility and more general theories of intentional control and automaticity.

Zusammenfassung

Eines der ungelösten Probleme menschlicher Kognition ist S. Monsell (1996) zufolge, wie genau sprachliche Aufgabeninstruktionen in Aufgabenrepräsentationen übersetzt werden, die instruiertes Verhalten steuern. Die vorliegende Arbeit versucht, Licht auf einen Aspekt dieser Frage zu werfen. Die spezifische Frage ist, ob und wie die Details der Instruktionen von Antworten in einfachen manuellen Zweifachwahlaufgaben die Kodierung und die Prozesse beeinflussen, die üblicherweise mit „Antwortselektion“ assoziiert werden, einem Verarbeitungsstadium, das als zentral für die willkürliche Steuerung von Handlungen angesehen wird. Thematisch liefert die Dissertation somit einen Beitrag zu der Frage nach den kognitiven Grundlagen der Steuerung von Willkürhandlungen.

Unter der Annahme, dass die spezifischen Inhalte sprachlicher Antwortinstruktionen die Antwortkodierung determinieren, ist zu erwarten, dass identische Aufgaben bei unterschiedlicher Antwortinstruktion unterschiedlich bearbeitet werden. Diese Vorhersage wurde mittels zweier experimenteller Ansätze in fünf Experimenten überprüft, in denen linke und rechte Tastendruck-Reaktionen entweder räumlich (als „linke“ vs. „rechte“ Taste) oder farblich (als „blaue“ vs. „grüne“ Taste) instruiert wurden. Es wurde untersucht, welchen Einfluss Antwortinstruktionen auf zwei Arten von Kompatibilitätseffekten haben.

In den ersten 3 Experimenten wurde ein Doppelaufgabenparadigma gewählt, das überlappende vs. nicht-überlappende Antworten auf einer manuellen und einer zeitgleich ausgeführten verbalen Aufgabe erforderte. Die verbale Aufgabe erforderte ebenfalls entweder „links“- und „rechts“- oder „blau“- und „grün“-Antworten. Wenn die Antworten beider Aufgaben räumlich (Experiment 1) oder farblich (Experiment 2) instruiert wurden, waren kompatible Antworten (z. B. verbale „blau“-Reaktionen gefolgt von blauen Tastenreaktionen) in beiden Aufgaben schneller als inkompatible. Wenn jedoch die verbale Aufgabe „links“- und „rechts“-Reaktionen verlangte, während die Tasten der manuellen Aufgabe farblich instruiert wurden, zeigten sich keine Kompatibilitätseffekte.

Das 4. und 5. Experiment dieser Arbeit erweitern die Doppelaufgabenexperimente dahingehend, dass der Einfluss der gleichen Antwortinstruktionsmanipulation auf den „Simon-Effekt“ (schnellere Antworten bei Korrespondenz als bei Inkorrespondenz zwischen Antwortposition und irrelevanter Stimulus-Position) mit Hilfe einer Aufgabe untersucht wurde, in der linke und rechte Tastenreaktionen willkürlich zentral dargebotenen Stimuli (Buchstabenidentität) zugeordnet wurden. Go/no-go Signale, die zufällig an unterschiedlichen Positionen

erschienen, gaben an, ob reagiert werden sollte oder nicht. Während ein Simon-Effekt bei räumlicher Antwortinstruktion in Experiment 4 beobachtet werden konnte, führten Farbinstruktionen der Antworttasten in Experiment 5 zu einer signifikanten Reduktion des Effekts.

Zusammengenommen legen diese Ergebnisse nahe, dass die in der Antwortinstruktion genutzten Antwort-„Label“ direkt bestimmen, welche Codes zur Reaktionssteuerung genutzt werden, und dass nicht-räumliche Antwortkodierung bei nicht-räumlicher Antwortinstruktion dominiert. Die Implikationen der Befunde für aktuelle Kodierungstheorien zur Erklärung von Kompatibilitätseffekten werden diskutiert und in Bezug gesetzt zu allgemeineren Theorien und Fragen zur willkürlichen Steuerung von Verhalten und zu den Bedingungen von Automatizität.

1 Introduction

Humans are able to seemingly effortlessly and rapidly translate relatively arbitrary instructions into behavior. Thus, in everyday life, most people are capable of installing a new Ikea BILLY book shelf without several days of frustrating trial and error installation attempts, provided they have carefully read the instructions and the IKEA set contains all parts and devices mentioned in the instructions. Similarly, volunteer participants arriving at a psychology lab are generally capable of following arbitrary task instructions almost immediately. Consider the following scenario. As soon as a research participant is seated in front of a computer screen I tell him or her something like the following: “On each trial, a word will be presented. Your task is to press the left key if the word refers to something alive (a person, an animal, or a plant; e.g., MOUSE), and to press the right key if the word describes an inanimate object (e.g., SHELF). When you have responded, the word will disappear and another trial will begin. There are 30 trials in each block and the first block is for practice. Ready?” The participant will probably answer “Uhm, er, yes, I suppose so”, possibly after first affirming that he or she really ‘got’ the correct mapping from stimulus categories to keys, and then the sequence of trials begins. The participant’s first responses will be a little hesitant. However, given the subject is willing to comply with the instructions, by the end of the first block, the participant is responding confidently and reasonably accurately.

In order to perform such an arbitrary task, people must somehow configure their cognitive system in a way that it “knows,” for example,

- which (of many possible) stimulus dimensions (e.g., animacy instead of the number of syllables) to focus on,
- how to map particular instances of stimuli to the relevant stimulus categories (e.g., animate vs. inanimate),
- how to map the stimulus alternatives to arbitrarily assigned responses (e.g., left and right keypresses),
- and how to access and execute these responses.

In short, people must adopt a ‘task set’ that implements an “effective intention to perform a particular task, regardless of which of the range of task-relevant stimuli will occur” (Rogers & Monsell, 1995, p. 207). Given that human beings, provided their brains are intact and mature (see, for instance, Luria, 1961), are able to follow arbitrary instructions, (verbal) instructions seem to somehow determine how task sets are configured. However, although most researchers would probably agree that experimental instructions are important for the outcome of an experiment, perhaps surprisingly, relatively little is known about how exactly task instructions are translated into, and are used to control behavior (cf. Monsell, 1996, who considers this one of the "unsolved mysteries of mind").

In this dissertation, I am concerned with simple binary stimulus-response instructions involving spatially organized keypress responses, such as, for example, "when you see a square on the screen, then press the left key; when you see a circle on the screen, then press the right key." My focus will be on how the specific contents of instructions affect the codes and processes commonly associated with response selection, a processing 'stage' assumed to be central in action control. My main question of interest is in whether or not the specific response labels given in such task instructions (e.g., "left" and "right" vs. "blue" and "green") play any role in the control of the instructed behavior. That is, whether variations in response instructions (e.g., instructing response keys in terms of location vs. color) affect how responses are accessed, and hence how an otherwise identical task is performed.

As outlined in Chapter 2.1, there are at least two possible theoretical positions although general theories of action control (e.g., Cohen, Braver, & O'Reilly, 2000; Logan & Gordon, 2001) remain rather vague with respect to this question. On the one hand, task instructions might set up general constraints on how actions can be coded in order to meet task demands. According to this view, termed 'constraint hypothesis', the response labels used in the instruction do not directly determine response coding. Rather, responses are coded in terms of features that allow to discriminate between possible response alternatives in the context of any given task instruction.

On the other hand, it is also conceivable that instructed response labels directly influence response coding. For example, a simple stimulus-response instruction might set up a link between the stimulus and the response components of the instruction by activating and linking the corresponding concepts (categories) mentioned in the instruction. According to this view, the 'direct coding hypothesis', instructed response codes become included in the response representations and can be used to control responding.

In Chapter 2.2, these broad theoretical positions will be elaborated with respect to spatially organized keypress responses. To this end, I will discuss the assumptions inherent in a subclass of coding accounts (i.e., contemporary dual route models) that have been proposed to explain so-called compatibility effects.

‘Compatibility effects’ refer to variations in reaction time and accuracy that occur as a function of the way in which (a) stimuli are assigned to responses (stimulus-response-compatibility; e.g., faster left hand responses to left pointing than to right pointing arrows), (b) responses on two concurrently performed tasks are paired (response-response compatibility; e.g., faster responses on two simultaneously performed tasks when both tasks require ‘left’ responses than when one task requires a ‘left’ and the other a ‘right’ response), or (c) response effects that appear contingent upon responding are assigned to responses (response-effect compatibility; e.g., faster left responses when ‘left’ rather than ‘right’ stimuli are presented upon responding).

Such compatibility effects are typically attributed to the ‘response selection’ stage. That is, most accounts of such effects assume that some sort of match between (features of) the response representations, on the one hand, and (features of) the stimuli, anticipated response effects, or responses on a simultaneously performed task, on the other hand, leads to automatic priming of the corresponding response, which is beneficial when the correct response is primed, but leads to response competition when this is not the case. In Chapter 2.2, three classes of such accounts are distinguished that differ with respect to their assumptions on how response instructions influence the coding of spatially organized keypress responses, and hence make different predictions regarding which match or compatibility relations should lead to compatibility effects under which instruction conditions.

One class of models holds a strong ‘spatial is special’ view. According to this position, termed ‘spatial coding’ hypothesis, spatially organized responses are coded in terms of (left-right) response location whenever this dimension allows discrimination of responses (e.g., Heister, Schroeder-Heister, & Ehrenstein, 1990; Lu, 1997; Roswarski & Proctor, 2003a). Because the spatial coding hypothesis assumes instruction-independent spatial response coding, this view can be considered to represent the more general constraint hypothesis outlined above where spatially organized keypress responses are concerned.

In contrast, other accounts seem to hold a ‘direct coding’ view by proposing that instructed response codes become included into the response representations and can be used to control responding even when the instructed response-dimension is non-spatial. I will argue that two versions of such a direct coding view can be distinguished.

One version, as, for instance, represented by the dimensional overlap model (e.g., Kornblum, Hasbroucq, & Osman, 1990) assumes that both instructed and uninstructed response codes are included in the response representations and equally contribute to responding. That is, instructed (non-spatial) response codes cannot be weighed more strongly than 'default' spatial response codes. Because this view implies restricted top-down control of response coding it can be considered a weak version of the direct coding hypothesis.

In contrast, according to the strong version of this view, the specific motor programs (or motor codes) that are needed to perform the instructed response might primarily be accessible via the mental representation activated by the response label. Such a view is consistent with the 'intentional feature weighing hypothesis' that was recently proposed by Hommel and colleagues (Hommel, Müsseler, Aschersleben, & Prinz, 2001). According to the intentional weighing hypothesis, instructed (intended) response features (that can be relatively abstract) are weighed more strongly than are irrelevant features, although the latter may still be part of the action representations.

In Chapter 3, I will review evidence from the compatibility literature involving spatially organized keypress responses that is consistent viz. inconsistent with the spatial coding hypothesis and the weak and strong versions of the direct coding hypothesis. The focus will be on two broad classes of compatibility effects that bear most directly on the experiments presented in the empirical part of this thesis. The first class (Chapter 3.1) is concerned with a variety of stimulus-response compatibility effects. The second class of compatibility effects to be reviewed (Chapter 3.2) are response-response compatibility effects obtained in dual task studies that require consistent viz. inconsistent responses on the two tasks.

The main question throughout this literature review will be whether variations of response instructions affect how a task is performed, that is, whether or not instructed response labels determine how responses are coded and accessed. If participants code their responses as instructed, one would expect that compatibility effects can be observed with respect to the instructed dimension even when the instructed response dimension is non-spatial and the response-overlapping stimulus- (or concurrent response-) attribute is task irrelevant. Accordingly, such findings (e.g., an impact of irrelevant stimulus color on responding when responses are instructed in terms of color) are interpreted as evidence in favor of the direct coding hypothesis. Moreover, if participants are able to weigh response codes as instructed, then response instructions that do not refer to the spatial dimension should affect the size or even the direction of spatial compatibility effects under consideration. Accordingly, such findings are interpreted as supporting the strong version of the direct coding hypothesis (i.e., the intentional weighing hypothesis).

In contrast, instruction independent spatial effects and a lack of irrelevant effects for other than spatial instructed response dimensions are considered more consistent with the spatial coding hypothesis.

The goal of the experiments presented in the empirical part of the thesis (Chapters 4 and 5) was to extend existing findings and to assess more directly as has been done before inhowfar the response labels used in the verbal task instructions determine response coding, and hence, performance. The general logic underlying the experiments was to vary response instructions for manual (left and right) keypress responses to arbitrary stimulus attributes. This was done by instructing the response keys as either left vs. right keys (spatial instructions) or as blue vs. green keys (color instructions). If participants arbitrarily code and access their responses as instructed, then response instructions should determine how responding is controlled. I used two different experimental approaches to address this general prediction, both relying on the compatibility logic outlined above.

In one set of experiments (Experiments 1-3, Chapter 4), a dual task methodology similar to that used by Hommel (1998, Experiment 1) was employed. More specifically, in addition to a manual keypress task with varied response instructions, participants had to perform a verbal task that either required “left” vs. “right” or “blue” vs. “green” concurrent verbalizations. When responses on the two tasks were both instructed in terms of location (Experiment 1) or color (Experiment 2), then compatible responses on the two tasks were faster than incompatible responses. However, when the verbal task required “left” vs. “right” responses whereas manual keypresses were instructed as blue vs. green (Experiment 3), then no compatibility effects were observed. These results suggest that response labels used in the instruction determine the codes that are used to control responding, hence supporting the strong version of the direct coding hypothesis.

Experiments 4-5 (Chapter 5) extend these results by employing the same response-instruction logic to a Simon-like task, in which left and right keypress responses were arbitrarily mapped to centrally presented stimuli (letter identity). Go/no-go signals presented at randomly varying locations indicated whether the prepared response was to be executed or not. Color instructions of the response keys (Experiment 5) significantly reduced the Simon effect (i.e., faster responses when response location and irrelevant go/no-go position correspond) observed under spatial response instructions (Experiment 4). This result seems to indicate that spatial response coding is a prerequisite for the Simon effect to occur, and, more importantly, it corroborates the findings from the first set of experiments that response instructions at least partially determine whether responses are spatially coded.

2 Instructions and Response Coding: Theoretical Positions

Research participants generally only respond when they are asked to do so (or when they infer that they are supposed to do something). They do not usually produce a response simply because they have registered some stimulus (e.g., a triangle). Models of action control frequently assume that, upon instruction, task sets are formed that specify the relevant stimulus dimension, the required responses, and the task-relevant mappings from stimuli to responses. Task sets can thus be viewed as behaviorally relevant representations of the to-be-performed task, implementing the goal to respond to certain (classes of) stimuli in a specific way. In the first section of this chapter (Section 2.1) I sample from recent theories of action control (i.e., Cohen et al., 2000; Logan & Gordon, 2001) and describe their assumptions on how instructions are translated into effective task sets. Although their predictions regarding the impact of specific response labels on action control remain relatively vague, they are consistent with at least two general theoretical positions.

In the second section of this chapter (Section 2.2) these positions will be elaborated and specified with respect to spatially organized keypress responses (i.e., left and right keypresses), on which the emphasis in the remainder of this thesis will be. To this end, I will provide a review of dual-route models of response selection (or more precisely, response activation) that have been proposed to explain so-called compatibility effects.

2.1 Task Representations and the Control of Action

Current models of action control tend to assume that behavior is controlled at different 'levels'. Furthermore, it is commonly believed that verbal instructions (or the internal representations of the language input) do not directly control behavior, but that instead verbal instructions have to be encoded/translated/compiled into other types of internal representations that, in turn, control behavior.

For example, the prefrontal cortex (PFC) model by Cohen et al. (2000; see also O'Reilly, Braver, & Cohen, 1999), well-known implementations of which have been provided by Jonathan Cohen's interactive activation models of Stroop performance (e.g., Cohen, Dunbar, & McClelland, 1990), assumes that instructions are encoded into discrete, combinatorial, and self-maintaining PFC representations containing internal contextual information. While

O'Reilly et al. (1999; see also O'Reilly & Soto, 2002) concede a special role for the phonological loop in the encoding and maintenance of task relevant information, the resulting PFC representations themselves are not assumed to be verbal, albeit possibly symbolic and categorical. Moreover, according to the PFC model, it is not the PFC alone that controls behavior. Rather, according to the model, PFC representations are or can be used to bias and constrain the activation flow in another network, the perceptual-motor cortex (PMC) layer, that is characterized by highly distributed representations and slow integrative learning through inductive weight changes. Whereas the PFC (in some versions of the model assisted by fast-learning mechanisms such as hippocampal systems) is responsible for maintaining (instructed) task goals and for constraining the behavioral PMC network accordingly, it is the PMC layer that processes stimuli and generates responses.

Similarly, in the Logan and Gordon (2001) model that was designed to model executive processes and phenomena associated with dual-task performance, verbal instructions are parsed into propositional task level representations that are stored in working memory. However, it is not the propositional representations that control behavior. Instead the propositional representations of instructions need to be translated into a set of parameters (e.g., response categories and weights for biases, attentional weight parameters, etc.) at the parameter level, which, in turn, are passed down to 'lower level' behavioral modules responsible for stimulus identification and categorization as well as response selection. Thus, in the Logan and Gordon (2001) model, the parameter set that is passed down constitutes the 'effective' task set.

In short, both the PFC model and the Logan and Gordon model assume that behavior is controlled at different levels, and that instructions need to be translated into a 'language of the mind', that is, into a representational format that sets up internal constraints, which in turn allow effective control of behavior.

However, neither of these more general models contains a principled account of how exactly instructions are translated into effective task sets. For instance, in different (hybrid connectionist) implementations of the PFC model, the PFC representations contain information about either the relevant stimulus dimension, or the relevant position, or specific response alternatives (see Botwinick, Braver, Barch, Carter, & Cohen, 2001; Cohen et al., 2000). Hence, what is extracted from instructions seems to depend on task demands, that is, of what needs to be controlled or what researchers assume needs controlling on a given task.

With regard to my main question of interest, namely whether or not the response labels used in simple binary stimulus-response (S-R) instructions affect how a task is performed, the PFC model in its current form therefore seems consistent with two different general positions. On the one hand, it is conceivable that the bias exerted by PFC representations is rather unspecific with regard to response coding. That is, it might just set up general constraints (e.g., respond to color instead of color words), whereas specific S-R mappings, and hence responses, are coded within PMC in a way that allows to discriminate all possible responses in a certain task context by biasing (pre-) existing processing pathways.

On the other hand, however, it is also conceivable that PFC representations "mediate an appropriate behavioral response" (Cohen et al., 2000, p. 196) by exerting a more direct influence, which may well depend on the specific contents of instructions (e.g., respond "left"). In order to arrive at more specific predictions with respect to the influence of response labels on response coding, a more precise account of the nature of the assumed PFC and PMC representations and their susceptibility to instruction is needed[1].

Similarly, in order to derive unequivocal predictions regarding the impact of response instructions on response coding, the Logan and Gordon (2001) model needs to be more specific with respect to (a) how parameters are extracted from the (propositional) task level representation, and, (b) the nature of response representations. As is, the model assumes that the 'response set' consists of response-relevant (stimulus) categories (e.g., odd vs. even for number stimuli) that are mapped to response counters. The latter are incremented according to stimulus categorizations "that correspond to them" (p. 400). Furthermore, it is assumed that "a later motor stage [not covered by the model …] turns a symbolic representation of the response into an overt action" (p. 396). Consequently, the Logan and Gordon model is again open to two interpretations regarding response coding. According to one, responses are coded and accessed in terms of the categories they are supposed to signal (e.g., as meaning "odd"). According to this interpretation, response coding in the Logan and Gordon model would be similar to Meiran's (e.g., Meiran, 2000; 2001) notion of response recoding, and would directly depend on instructions (i.e., on the response relevant stimulus categories).

On the other hand, however, it is also conceivable that the 'counters' responsible for response selection are coded in an instruction-independent way (e.g., as left and right), and that

---

[1] It should be noted that the authors themselves acknowledge this shortcoming in their model by stating that they "have not yet specified what this [the representational scheme of the PFC] is, nor the principles that might characterize it," (Cohen et al., 2000, p. 207) and by declaring this a major goal for the future.

the categorization evidence is (automatically) transmitted to the responses assigned to them. In the latter case, response coding and access would be logically independent from stimulus categorization, and, possibly, response instruction.

In sum, both models seem consistent with two general theoretical positions concerning how specific response labels given in task instructions for manual two-choice tasks (e.g., “when you see a square, then press the blue key; when you see a circle, then press the green key”) affect response coding, and hence response selection.

On the one hand, task instructions might set up general constraints on how actions can be coded in order to meet task demands. According to this view, the response labels used in the instruction do not directly determine response coding. Rather, responses are coded in terms of features that allow to discriminate between response alternatives in the context of any given task instruction. In what follows, this view will be termed the ‘constraint hypothesis.’

On the other hand, however, it is also conceivable that instructed response labels directly influence response coding. For example, a simple S-R instruction might set up a link between the stimulus and the response components of the instruction by activating and linking the corresponding concepts (categories) mentioned in the instruction. The motor programs[2] that are needed to perform the instructed response might then also be accessible via the mental representation activated by the response label. In the following, this position will be termed the ‘direct coding hypothesis.’

Note that the different coding hypotheses are not necessarily mutually exclusive. For instance, the direct coding hypothesis does not preclude the possibility that instructed codes are merely added to some sort of ‘default’ representation of responses. In this case, it is conceivable that participants use instructed codes only in the beginning of working on a new task, or when considered useful.

In the next section, the two general theoretical positions will be discussed with respect to spatially organized keypress responses (i.e., left and right keypresses), on which the emphasis in the remainder of this thesis will be. To this end, I will interpret and classify so-called dual route models of stimulus-response-compatibility (SRC) with respect to their assumptions regarding response coding.

---

[2] Here and in the following, I use the terms ‘motor program’ and ‘motor code’ interchangeably.

2.2 Response Coding

Dual route models of compatibility have been proposed to explain SRC effects. SRC effects are variations in reaction time (RT) and accuracy that occur as a function of the way in which stimuli are assigned to responses. The general finding has been that responding is easier (faster and less error prone) in the compatible (matching) than in the incompatible condition. For instance, left keypress responses to stimuli appearing on the left of the screen (compatible or matching condition) are faster than left responses to right stimuli (incompatible condition) (e.g., Broadbent & Gregory, 1962). Such performance differences are observed even when irrelevant stimulus attributes match viz. mismatch the required response, as is the case, for instance, in Simon-type tasks (for reviews, see Lu & Proctor, 1995; Simon, 1990). In a typical Simon task, subjects are required to respond to arbitrary stimulus attributes such as pitch of a tone or the color of a visual stimulus by pressing spatially organized (usually left vs. right) keys, while stimulus position varies randomly. Although stimulus position is task irrelevant, it affects performance such that responses are faster (and more accurate) when stimulus position and the side of the required response correspond than when they do not correspond.

Dual-route models (e.g., Barber & O'Leary, 1997; De Jong, Liang, & Lauber, 1994; Hommel, 1997; Kornblum et al., 1990; Tagliabue, Zorzi, Umiltà, & Bassignani, 2000; Zhang, Zhang, & Kornblum, 1999) represent an influential subclass of coding accounts that have been proposed to explain such SRC effects, and appear to be particularly well suited to handle irrelevant SRC effects such as the Simon effect.

Different manifestations of dual route models share the assumption that response selection is affected by two, more or less, independent routes. One of the routes, alternatively labeled short-term memory (STM) link(s), indirect link(s), conditional or conditionally automatic route, or translation route, directly depends on instruction. In most models, this route (if explicitly modeled at all) is implemented by links that connect internal representations of the task-relevant stimulus attributes (e.g., codes of the letters A and B in Figure 1) to representations of the responses assigned to them (see solid arrows in Figure 1). Some models explicitly distinguish between stimulus feature codes and a hidden layer (see STM nodes in Figure 1) coding "task relevant attributes" (e.g., Tagliabue et al., 2000, p. 661) that mediate S-R translation in the STM route.

According to most of the newer dual route models, activation is transmitted automatically along these links once they are implemented. However, because these links depend on instructions (i.e., on task-relevant stimulus attributes and their assignment to responses), this route is considered conditionally automatic (e.g., De Jong et al., 1994).

In addition to the indirect route provided by the STM links, stimuli are assumed to activate their "corresponding" responses via a direct route (also called long-term memory (LTM) links, unconditional or unconditionally automatic route) if stimulus and response attributes (codes) overlap. Since activation along these direct links does not depend on the task relevance of the stimulus attribute that elicits it, this route is considered unconditionally automatic (see broken arrows in Figure 1).

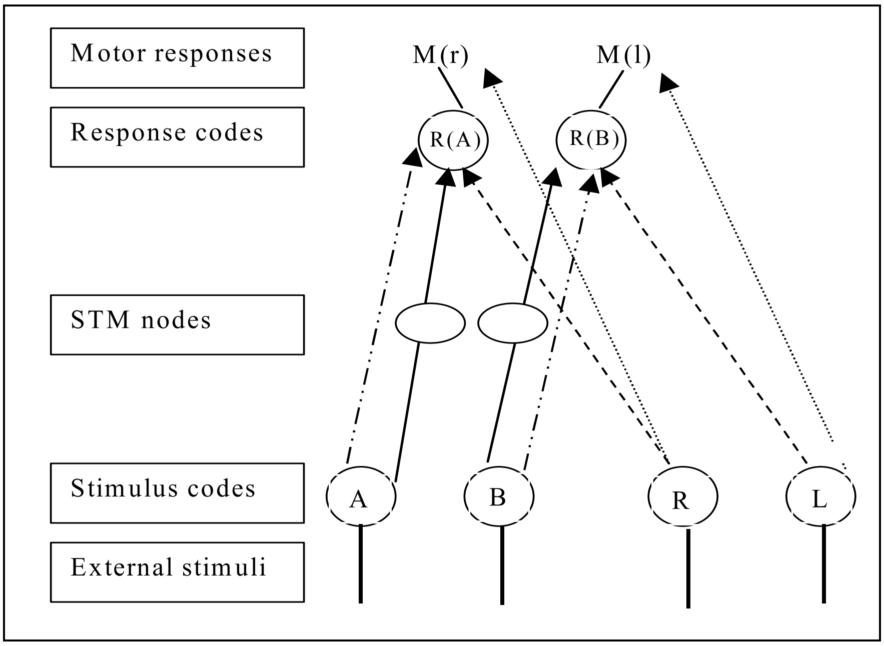

Figure 1. A schematic illustration of the core assumptions of (different classes) of dual route models. Solid arrows represent STM links that connect task relevant attributes (i.e., letter identity) with representations coding the required responses via STM nodes. The broken arrows indicate direct links connecting overlapping stimulus and response features, regardless of whether the respective stimulus attributes are task relevant or not (L=left and R=right stimulus position codes; R(A) and R(B): response representations linked to stimuli A and B; M(r) and M(l): right and left hand motor programs). See text for details.

Although I agree with Hommel (1996a, p.108) that "it is clear that a principled account of response coding is lacking," in most coding accounts of SRC, dual route models seem to allow inferences regarding their often implicit or vaguely formulated response coding assumptions. This is so because activation is only transmitted via the direct route if stimulus and response codes overlap. Thus, a closer inspection of what types of S-R overlap lead to direct activation of spatially organized responses (e.g., left and right keypress responses) in the unconditional route provides insights regarding how responses are assumed to be coded and accessed.

Roughly, two classes of models can be distinguished regarding their assumptions concerning direct response activation, and, by implication, how spatially organized responses are thought to be represented and accessed.

One class of models seems to hold a strong ‘spatial is special’ view, in that they assume that only spatial stimulus attributes (i.e., the L and R codes in Figure 1) directly (unconditionally) activate their respective responses, implying that responses are assumed to be spatially coded regardless of instructions. This assumption comes in three flavors.

The most widely accepted version of this view is represented by spatial coding accounts (e.g., Barber & O’Leary, 1997; Heister et al., 1990; Lien & Proctor, 2002; Lu & Proctor, 1995). It holds that both the indirect and the direct route converge on cognitive response codes that (primarily) represent relative key position (instead of the anatomical motor codes themselves; see the dashed lines in Figure 1 that connect stimulus position codes and response codes R(a) and R(b)), which in turn activate their corresponding motor responses. According to this view, responses are selected on the basis of spatial response codes representing relative key position whenever key position allows the discrimination of responses. A second, albeit less prominent (cf. Roswarski & Proctor, 2003b) version of this view is represented by motor priming accounts of the Simon effect (e.g., Wascher, Schatz, Kuder, & Verleger, 2001). They propose that (certain types of) spatial stimulus attributes directly specify the motor parameters of the required lateralized responses, without any intervening cognitive response codes (see dotted lines, Figure 1, that directly connect the stimulus position codes R and L with their corresponding motor programs M(r) and M(l)). Finally, according to a third version of the spatial view (e.g., De Jong et al., 1994; Tagliabue et al., 2000), stimulus position is unique (and the only source of direct activation) because of a “natural tendency to react toward the source of stimulation” (Simon, 1969, p. 174). The notion of such a ‘natural tendency’ seems rather unspecific, and, in principle, appears to be consistent with both, a location coding and a motor priming interpretation of spatial coding.

Although all versions of the ‘spatial is special’ view share the basic assumption that spatially organized responses such as bimanual keypress responses are somehow spatially coded whenever spatial coding allows discrimination of the responses, in what follows, I will restrict the term ‘spatial coding hypothesis’ to the first, location coding, view unless otherwise noted. According to this view, instructing response keys non-spatially (e.g., by using symbolic non-spatial response instructions and labels, such as response labels A and B, see Figure 1), does not directly affect response coding. Rather, if response instructions have any effect at all their impact is restricted to some intermediate translation stage in the conditional route. That is, they are assumed to only affect translation efficiency in the indirect route through usually ill-defined (stimulus and/or response) recoding processes[3] that facilitate or, in case of incompatible mappings, hinder translation from relevant stimulus attributes to the spatially coded responses (cf. De Jong et al., 1994; Lu & Proctor, 1995). This implies that so-called symbolic SRC effects, such as faster red-key responses to red than to green stimuli when responses are instructed in terms of color, are not attributed to a match between stimulus and <u>response</u> codes, but instead to a match viz. mismatch of codes (e.g., verbal labels) at some intermediate stage that leads to selection of spatially coded responses (e.g., Bashore, 1990; for a more recent explication, see Mattes, Leuthold, & Ulrich, 2002).

The spatial coding hypothesis can thus be considered representative of the constraint hypothesis (see Chapter 2.1) where spatially organized keypress responses are concerned: Instructions only set up general constraints on how the conditional route is configured, and hence, how relevant stimulus attributes are translated onto responses without affecting response coding per se. Rather, responses are coded and accessed in terms of relative location whenever the spatial dimension allows discrimination of responses.

In contrast, a second class of dual route models seems to be more consistent with the direct coding hypothesis, which holds that instructed response labels directly influence response coding, for instance, by priming the corresponding concepts (categories) that are integrated into the response representations, and that can subsequently be used in response selection. Proponents of these models explicitly (e.g., Hommel, 1997; Hommel et al., 2001; Kornblum, Stevens, Requin, & Whipple, 1999) or implicitly (e.g., Kornblum et al., 1990; Zhang et al., 1999) assume that responses are represented such that every response is coded in terms of its features, dimensions, or categories (e.g., as being blue, left, manual, leading to a high pitch response effect etc.). Stimuli that overlap with respect to any of these features are assumed to (unconditionally) automatically activate their corresponding response features, regardless of whether this feature or dimension is spatial or not, and whether the overlapping stimulus is

---

[3] Typically, differential translation efficiency for different types of mapping has been modeled by simply assigning either high or low weights to the links from STM nodes to response codes (e.g., Tagliabue et al., 2000).

task relevant or not. Hence, according to this view, the direct route is not restricted to the spatial dimension, but extends to all overlapping features (see dash-dot-dot lines from stimulus nodes A and B to response keys labeled A and B in Figure 1). As a consequence, these models do not make a principled distinction between spatial and symbolic compatibility effects.

Within this second class of dual route models, two positions can be distinguished that differ regarding their assumptions concerning intentional weighing of response codes, and, by implication, with respect to the role of spatial response codes under non-spatial response instructions.

The weak version of the direct coding view, as, for example, represented by all instantiations of the dimensional overlap (DO) model proposed by Kornblum and colleagues (e.g., Kornblum et al., 1990; Kornblum et al., 1999; Zhang et al., 1999) does not distinguish between overlap on 'implicit' (uninstructed) and 'explicit' (instructed) dimensions. That is, the strength of direct S-R activation is the same regardless of whether a stimulus attribute is task relevant or not, and, more importantly, whether certain response features are task relevant (e.g., instructed) or not. This view is, for instance, reflected in Zhang et al.'s (1999) implementation of a task in which colored stimuli that randomly appear to the left or right of fixation are mapped to left and right keypress responses that are instructed (and labeled) in terms of color (i.e., the Hedge & Marsh (1975) task, see Chapter 3.1.4). In their model of this task, stimulus position codes and stimulus color codes (directly) activate response codes to the same extent. Thus, uninstructed spatial default codes are not only assumed to be part of the response representations under non-spatial response instructions, but they are weighed as strongly as instructed (non-spatial) response categories.

In contrast, Hommel et al. (2001; also see Hommel, 1997) assume that both stimulus and response features can be differentially weighed, depending on task demands. More specifically, a core assumption of the theory of event coding is that action representations include codes of (perceivable) proximal and distal action effects (e.g., a "left" proprioceptive feedback, a loud click on the left side, a light on the right that is turned on by a left keypress and so on). According to the theory of event coding, responses are accessed via their intended (anticipated) action effects. Direct S-R activation occurs in this model as a consequence of overlap of features that are used to code both stimuli and responses in a common representational medium. As in the DO model, feature overlap, and hence, direct activation, is therefore not restricted to the spatial dimension. Importantly though, the theory of event coding assumes that features can be weighed according to task demands (e.g., instructions). Hence, intended action effects contribute more strongly to response coding and response control than implicit (non-intended) features, although the latter may still be part of the response representation. As a consequence, compatibility arising from overlap between (irrelevant) stimuli and instructed (intended) response codes can be expected to override ‘implicit’ S-R overlap. Because instructed coding can dominate uninstructed coding, this view can be considered a strong version of the direct coding hypothesis.

Taken together, two broad theoretical positions have been identified with regard to the main question of this thesis, namely whether or not the specific response labels given in simple binary choice task instructions involving spatially organized keypress responses determine how such a task is performed, that is, how responses are coded and selected. According to the spatial coding hypothesis (e.g., De Jong et al., 1994; Lu, 1997; Roswarski & Proctor, 2003a), which represents the constraint hypothesis regarding spatially organized keypress responses, response labels used in the instruction do not directly determine response coding. Rather, responses are coded in terms of relative key location whenever the spatial dimension allows discriminating between responses.

On the other hand, the direct coding hypothesis assumes that instructed response labels directly influence response coding by activating and linking the corresponding concepts (categories) mentioned in the instruction. According to both the weak and the strong versions of the direct coding hypothesis, instructed codes are included into the response representations and contribute to response selection. However, not even proponents of the direct coding hypothesis propose that non-spatial response instructions lead to complete substitution or elimination of spatial response codes. Rather, they assume that spatial (default) codes are part of the response representation and are or can be used to access responses even when response instructions do not refer to the spatial dimension. Whereas the <u>weak</u> direct coding hypothesis holds that spatial response coding is largely unaffected by the inclusion of instructed (non-spatial) response codes, the <u>strong</u> direct coding hypothesis assumes that response codes can be weighed according to instruction. Consequently, only the strong direct coding hypothesis predicts that instructed (non-spatial) codes or features can dominate spatial coding.

In Chapter 3, I will review evidence from the compatibility literature involving tasks requiring spatially organized keypress responses that is consistent viz. inconsistent with the spatial coding hypothesis and the strong and weak versions of the direct coding hypothesis.

The main question throughout this literature review will be whether response instructions affect the size or the direction of the compatibility effects under consideration.

3 Empirical Evidence for and against Direct Coding

As outlined above, compatibility effects are commonly attributed to a stage in information processing known as the ‘response selection stage’. More specifically, the assumption is that some sort of match between (features of) the response representations that are used to control responses on the one hand, and (features of) the stimuli (→ stimulus-response compatibility), response effects (→ response-effect compatibility), or responses on a simultaneously performed task (→ response-response compatibility), on the other hand, leads to priming of the corresponding response, which is beneficial when the correct response is primed, but leads to response competition when this is not the case. According to this logic, investigating which match or compatibility relations contribute to compatibility effects under which instruction conditions allows conclusions about the cognitive codes that are used to control instructed responding.

Therefore, this chapter provides a review of findings regarding the impact of response instructions on several compatibility effects involving spatially organized keypress responses. The focus will be on two broad classes of compatibility effects that bear most directly on the experiments presented in Chapter 4 and Chapter 5.

The first class (Chapter 3.1) deals with a variety of SRC effects. The SRC effects reviewed include (i) the left-right prevalence effect observed with two-dimensional spatial S-R mappings (Section 3.1.1), and (ii) spatial and non-spatial SRC effects observed when irrelevant stimulus attributes overlap with (features of) spatially organized responses, that is, variations of the Simon effect and the manual Stroop effect (see below, sections 3.1.2-3.1.4). Whereas instruction manipulations in the first two sections (i.e., Section 3.1.1-3.1.2) mainly involve an emphasis on different spatial aspects of the response array, the majority of findings reviewed in Sections 3.1.3 and 3.1.4 are concerned with non-spatial response instructions and response coding.

The second class of compatibility effects to be reviewed (Section 3.2) involves intertask consistency effects obtained in dual task studies that require compatible viz. incompatible responses on two concurrently performed tasks.

The main question throughout this literature review will be whether variations of response instructions affect the type, size, or direction of observed compatibility effects. According to both versions of the direct coding hypothesis, it should be expected that variations in response instructions affect response coding, and hence, how responses are accessed. Consequently, one would expect that compatibility effects can be observed with respect to the instructed dimension even when the overlapping stimulus- (or concurrent response-) attribute is task irrelevant. It will be argued that especially irrelevant non-spatial effects (i.e., symbolic Simon-type effects) provide convincing evidence for the direct coding hypothesis because they cannot (easily) be explained in terms of translation efficiency. The strong version of the direct coding hypothesis furthermore predicts that instructed codes are weighed more strongly than uninstructed codes. Accordingly, instruction manipulations (e.g., non-spatial vs. spatial response instructions) should lead to variations in the direction or size of a given spatial compatibility effect.

In contrast, the spatial coding hypothesis predicts instruction independent spatial compatibility effects (especially Simon-type effects) and a lack of irrelevant effects for instructed non-spatial response dimensions.

3.1 Response Instructions and Stimulus-Response Compatibility

As described in Chapter 2, stimulus-response compatibility refers to systematic variations in choice RT and error likelihood that depend on the relations between stimulus and response sets. Reaction times are shorter when there is correspondence between a stimulus attribute and features of the response (representation) than when there is not. In the following, I will review SRC phenomena obtained with spatially organized choice reactions (mostly bilateral keypress responses) that have been subject to instruction manipulations.

3.1.1 Right-Left Prevalence in Two-Dimensional Stimulus-Response Mappings

Right-left prevalence refers to the finding, first reported by Nicoletti and Umiltà (1984, 1985; Nicoletti, Umiltà, Tressoldi, & Marzi, 1988; for a summary, see Umiltà & Nicoletti, 1990), that under situations where stimuli and responses simultaneously overlap on the horizontal and the vertical dimension, compatibility on the horizontal dimension dominates vertical compatibility, regardless of instructions.

The prototypical task used to investigate this effect is the two-dimensional spatial mapping task. In this task, both horizontal and vertical compatibility are varied orthogonally such that a particular response is vertically and/or horizontally compatible or incompatible with the stimulus. For example, in one subtask, keys on the top-left and the bottom-right have to be pressed in response to stimuli on the top-left and bottom-right side respectively, yielding horizontal and vertical compatibility. In another subtask, the same keys have to be pressed in response to stimuli on the top-right and bottom-left respectively, yielding vertical compatibility and horizontal incompatibility (see Table 1 for examples of conditions in a paradigmatic experiment).

Instructions in this task have been varied by emphasizing either the vertical or the horizontal dimension. Vertical instructions refer to stimuli as well as responses only in terms of their vertical position (e.g., “press the upper key in response to the bottom stimulus”), without mentioning the horizontal dimension. Conversely, horizontal instructions only refer to the horizontal dimension (e.g., “press the left key in response to a left stimulus”).

Table 1. Sample stimuli and responses, and the resulting compatibility relations in four subtasks of the two-dimensional spatial mapping task.

Task/	Stimulus 1	Response 1	Stimulus 2	Response 2	Compatibility	
Block					Vertical	Horizontal
1	Top left	Top left	Bottom right	Bottom right	+	+
2	Top right	Top left	Bottom left	Bottom right	+	-
3	Bottom left	Top left	Top right	Bottom right	-	+
4	Bottom right	Top left	Top left	Bottom right	-	-

The weak version of the direct coding hypothesis does not discriminate between implicit and explicit overlap. Accordingly, it predicts that responses are coded with respect to both dimensions, without the possibility to weigh vertical and horizontal response codes. Therefore, symmetrical horizontal and vertical effects, both on the instructed dimensions (i.e., vertical effects under vertical instructions and horizontal effects under horizontal instructions) and the “irrelevant” dimensions (i.e., vertical effects under horizontal instructions and horizontal effects under vertical instructions) should be expected, although the effects on the instructed dimensions may be larger than on the uninstructed dimensions. This is so because both the conditional and the direct route can be expected to contribute to the former, but only the direct route would be responsible for the latter effect.

In contrast, the strong version of the direct coding hypothesis predicts that response codes can be weighed and used as instructed. Consequently, symmetric vertical and horizontal instructed effects (i.e., vertical | vertical and horizontal | horizontal) should be observed. Uninstructed (irrelevant) effects (i.e., vertical | horizontal and horizontal | vertical) should be negligible for both dimensions because the response codes corresponding to the uninstructed dimension can be assumed to be less strongly weighed.

The predictions of the spatial coding hypothesis are more ambiguous. This is so because none of the dual route models (see Chapter 2.2) adhering to the spatial coding view explicitly distinguish between the vertical and the horizontal spatial dimension. Because both dimensions are spatial, the spatial coding hypothesis is therefore consistent with the view that both dimensions are used for coding. In this case, the predictions of the spatial coding hypothesis would be identical with those of the weak version of the direct coding hypothesis (see above).

On the other hand, in a strong interpretation of spatial coding as satisfying constraint setting, which holds that responses are coded in terms of features that allow to discriminate all possible responses in a given task context, one dimension is sufficient for response coding. In the interpretation of spatial coding favored here, only one dimension (e.g., the left/right dimension) is used for response coding. Generally, one would therefore expect compatibility effects primarily for the 'default' dimension (e.g., the horizontal dimension), independently of instructions. However, in this task there is overlap regarding both the instructed and the uninstructed dimensions. Therefore, one might expect translation-based compatibility effects for the other (e.g., vertical) dimension as well when this dimension is task relevant (e.g., under vertical instructions), but not when the default dimension is instructed (e.g., under horizontal instructions). In contrast, a compatibility effect for the default dimension (e.g., horizontal effects) should be observed under both horizontal and vertical instructions, although the latter might be somewhat smaller. Because only the conditional route leads to vertical effects, but both the direct and conditional route can be expected to contribute to horizontal effects, the horizontal effect should be generally larger than the vertical effect.

The majority of results obtained with this task generally seem to support the spatial coding hypothesis (in its one-dimensional interpretation). That is, most of the studies that included both, a vertical and a horizontal instruction condition, and that required bimanual responses found what Vu, Proctor, and Pick (2000) termed 'weak horizontal prevalence' (e.g., Hommel, 1996a, Exp. 1A; Vu & Proctor, 2001, Exp. 2). That is, a substantial vertical compatibility effect shows up under vertical instructions, although the vertical effect under vertical instructions is usually smaller than the horizontal compatibility effect under horizontal instructions. Moreover, horizontal effects tend to be obtained under vertical instructions, whereas most often no vertical effect is observed under horizontal instructions. Therefore, the overall horizontal effect is typically larger than the vertical effect. The horizontal effect is usually also modified by instructions such that more pronounced horizontal effects can be observed under horizontal than under vertical instructions, possibly indicating the additional contribution of the conditional route under horizontal instructions.

These findings suggest that instructions modify behavior, but that they are less effective in influencing (i.e., overriding) left-right coding than top-bottom coding, at least when stimuli and/or bilateral responses can easily be discriminated on the left-right dimension (see Vu & Proctor, 2001, 2002, for manipulations of the salience, that is, discriminability, of the two dimensions). They contradict both versions of the direct coding hypothesis according to which symmetric effects would have been expected.

However, Hommel (1996a) noted that horizontal prevalence effects are open for alternative explanations, namely (a) recoding of instructions, and (b) logical recoding.

The first of these explanations is based on the observation that, in the two-dimensional mapping task, the correct response is always redundantly signaled by both spatial stimulus attributes. That is, within a given subtask, the correlation between each stimulus and response dimension is either +1 or -1 (see Table 1). Hence, subjects may have ignored vertical instructions in some cases, basing their responses on horizontal stimulus codes instead. This explanation seems reasonable given that some studies by Nicoletti and colleagues (e.g., Nicoletti & Umiltà, 1984, Exp. 4; 1985, Exp. 1; also see Vu & Proctor, 2001, Exp. 3, normal instructions) reported (a) larger horizontal than vertical effects under vertical instructions (in fact, vertical effects were small to nonexistent), and (b) overall RTs under vertical instructions that closely resembled RTs under a purely (one-dimensional) horizontal mapping rather than those obtained for a purely vertical mapping. Furthermore, when Vu et al. (2000, Exp. 4) compared normal vertical instructions with exclusion instructions that encouraged subjects to exclusively rely on the vertical dimension, the horizontal | vertical effect of 57 ms under normal vertical instructions was reduced to 8 ms under exclusion instructions, whereas the vertical | vertical effects were 19 ms and 79 ms under normal and exclusion instructions, respectively (no horizontal instruction conditions were included). This finding presumably implies that participants were more likely to use vertical codes under vertical exclusion instructions (but see Hommel, 1996a, Exp. 1A, for slightly different findings; Vu et al., 2000, for an alternative interpretation).

Further evidence for participants' use of some sort of recoding strategy – albeit a somewhat different one – stems from underadditive interactions between horizontal and vertical compatibility sometimes observed under vertical instructions (e.g., Hommel, 1996a, Exp. 1A; Vu et al., 2000, Exp. 1A). That is, with vertical incompatibility, horizontal incompatibility often leads to faster responses than horizontal compatibility, indicating that the benefit of spatial S-R compatibility decreases, and can even be inverted when there is incompatibility on the other dimension.

As Hommel (1996a) noted, this finding is similar to an interaction of spatial and non-spatial dimensions first observed in a study by Hedge and Marsh (1975; see Chapter 3.1.4 below). Hedge and Marsh's participants always responded to colored (red vs. green) stimuli that randomly appeared on the left or the right side of the screen by pressing lateralized response keys that were labeled by colors (i.e., red vs. green). In the direct mapping condition, participants were instructed to press the key of the corresponding color (e.g., the red key in response to a red stimulus), whereas the reversed mapping required opposite color responses (e.g., green key to red stimulus). Although the spatial dimension was irrelevant, a spatial effect was observed for both mappings. However, whereas responses were faster when stimulus position and response position overlapped (i.e., in the spatially compatible condition) with the direct mapping, under the reversed mapping spatially compatible responses were <u>slower</u> than incompatible responses.

One influential account for this crossover effect is the logical recoding account (e.g., De Jong et al., 1994; Hedge & Marsh, 1975). The basic tenet of this account (in one of its interpretations) is that participants recode the relevant and – inadvertently – also the irrelevant (spatial) stimulus attribute in order to form the response code. Under the direct mapping, so the assumption, some sort of "same" operation is applied to both stimulus dimensions, leading to faster responses when both color and position correspond with the required response (e.g., applying a "same" operation to a red, left stimulus leads to faster responses when the red key is located on the left). In contrast, under the reversed mapping condition a "respond opposite" rule is formed that transforms the values on both stimulus dimensions into their opposite. Consequently, a left stimulus attribute now primes a right response, leading to faster responses in the spatially incompatible than the compatible condition.

Similarly, under vertical instructions, a 'same' transformation may be applied in vertically compatible subtasks that leave the horizontal stimulus attributes unchanged. With vertical incompatibility however, participants may apply a 'reversal' transformation to both vertical and horizontal stimulus position, leading to slower responses on horizontally compatible than incompatible subtasks.

Thus, it seems possible that the right-left prevalence effect is at least partially due to a strategic bias that either leads people to ignore instructions (i.e., use of the left-right stimulus and/or response dimension only), or to logical recoding of stimuli under vertical instructions. Moreover, because the relevant stimulus attribute in these studies is always position, and because the irrelevant stimulus attribute is correlated with response location, it is impossible to determine whether this bias primarily refers to stimulus or response coding, or to some sort of interaction of strategically biased stimulus and response processing. Evidence against a response coding bias has been provided by studies that observed vertical compatibility effects of typical size (i.e., comparable to horizontal effects) in one-dimensional vertical mappings, both when stimulus position was task-relevant (e.g., Vu et al., 2000, Exp. 1B) and, more importantly, when stimulus position was task-irrelevant (i.e., in vertical Simon tasks, e.g., Stürmer, Leuthold, Soetens, Schröter, & Sommer, 2002, Exp. 1), suggesting that, in principle, left and right responses (i.e., effectors) can be coded as up and down.

Similarly, Proctor, Vu, and Nicoletti (2002) recently reported symmetric horizontal and vertical compatibility effects when both vertical and horizontal stimulus position were task irrelevant in a two-dimensional version of the Hedge and Marsh task. In the Proctor et al. (2002) study, colored stimuli that randomly appeared at the upper left and lower right or lower left and upper right side of the screen were assigned to colored response keys of the same color (direct mapping condition) that were again defined - but not instructed - by the two spatial dimensions. This finding again questions an interpretation of horizontal prevalence in the two-dimensional mapping task in terms of obligatory left-right coding of responses.

Taken together, the (weak) horizontal prevalence effect obtained in most studies using the two-dimensional mapping task seems generally more consistent with the spatial coding hypothesis in its one-dimensional interpretation than with both versions of the direct coding hypothesis. As would be expected if both the direct and the conditional route contributed to horizontal effects, but only the conditional route were responsible for vertical effects, horizontal effects under vertical instructions tend to be larger than vertical effects under vertical instructions, and horizontal effects typically occur even under vertical instructions, whereas no irrelevant vertical effects show up.

However, the sometimes observed interaction between vertical and horizontal compatibility as well as the fact that both stimulus dimensions signal the correct response in this task, suggest that the prevalence effect might at least in part be due to a strategic bias to recode vertical instructions (i.e., mappings). The Proctor et al. (2002) finding of symmetric horizontal and vertical Simon effects when both dimensions are task irrelevant (i.e., when mappings are instructed in terms of color) furthermore questions an interpretation of horizontal prevalence as indicating a bias in response coding. More likely, such a bias refers to stimulus coding or some sort of interaction of (strategic) stimulus and response processing.

Therefore, a stronger test of an underlying response coding bias would be provided by experiments in which a non-spatial stimulus attribute (e.g., color) were task relevant and top-bottom as well as right-left spatial stimulus positions varied randomly. If the horizontal compatibility effect under horizontal response instructions (e.g., press the left vs. right key) and across instruction conditions would still be larger than the vertical compatibility effect under both vertical response instructions (e.g., press the top vs. bottom key) and overall, then this would provide strong evidence for obligatory left-right (spatial) coding. In contrast, if the effects were symmetric and modulated by instruction, then this would provide evidence for the strong version of the direct coding hypothesis. Finally, according to the weak direct coding hypothesis, symmetric vertical and horizontal Simon effects should be observed in such an experiment, regardless of response instructions. To my knowledge, such a study does not yet exist, though.

In the next section, I will review a recent debate on the impact of response instructions on a variant of the Simon task that requires responding with hands in the standard position as compared to a condition with crossed hands. As in the two-dimensional mapping task, different instructions emphasize different spatial aspects of the response 'array'. However, in the Simon task, stimulus position is uncorrelated with response location, rendering stimulus position truly task irrelevant.

3.1.2 Stimulus-Hand Correspondence and the Simon Effect

In Simon-type tasks non-spatial stimulus attributes such as pitch of tone or the shape of a visual stimulus are mapped to spatially organized responses. Stimulus position varies randomly from trial to trial, and hence is irrelevant for selecting the appropriate responses (usually bilateral keypresses). Nevertheless, the usual finding – known as the Simon effect – is that responding is faster when stimulus and response position correspond. Typically, left and right keys are pressed with left and right effectors, respectively. Thus, effector and key position (and effector position) correspond, implying that this task cannot discriminate between location (i.e., spatial) and effector (i.e., anatomical) coding. This confounding can be avoided by requiring participants to respond with crossed hands, such that the left key is operated by the right hand, and the right key is operated by the left hand (see Table 2 for stimulus-response compatibility conditions as a function of hand position). Thus, comparing normal and crossed hands conditions allows conclusions about whether responses are (primarily) coded spatially or anatomically.

Table 2. Stimulus-response compatibility conditions realized by the tasks with uncrossed vs. crossed hands.

Mapping condition	Stimulus-response position compatibility	
	compatible	incompatible
uncrossed	+/+	**-/-**
crossed	+/-	-/+

Note. + and - indicate compatibility and incompatibility between stimulus position and response position / response hand, respectively.

Deriving specific predictions from the hypotheses and models outlined in Chapter 2 is difficult because even coding accounts of the first generation (e.g., Wallace, 1971), the more so the dual route models reviewed in Chapter 2, are based on the results typically obtained with the crossed-hands task.

That is, a location based Simon effect is usually observed even when hands are crossed. More specifically, the usual finding with Simon tasks comparing standard vs. crossed hands responses (e.g., Roswarski & Proctor, 2000, 2003a; Simon, Hinrichs, & Craft, 1970; Wallace, 1971) is that responses are fastest for the position-compatible uncrossed condition (see +/+ condition in Table 2) in which correspondence exists for all three compatibility relations (i.e., stimulus-key, stimulus-hand, and hand-key), and slowest for the position-incompatible crossed condition (see -/+ condition in Table 2) in which neither stimulus and key position nor key position and hand (but stimulus position and hand) correspond, leading to overall faster responses with uncrossed as compared to crossed hands. More importantly, location-based Simon effects of comparable size usually show under both uncrossed (i.e., +/+ vs. -/- comparison) and crossed (i.e., +/- vs. -/+ comparison) hands conditions. Grouped according to stimulus-hand compatibility and stimulus-response location compatibility (without including hand condition as a factor), the same pattern of results can also be described as showing a main effect of position compatibility (i.e., +/+ and +/- vs. -/+ and -/-) that is qualified by an interaction of position compatibility and stimulus-hand compatibility. Usually, no overall stimulus-hand compatibility main effect is observed.

This pattern of finding led most researchers to propose that responses are primarily spatially coded, without denying some (minor) impact (on the size of the effect, but not on its direction) of anatomical coding. More specifically, most coding accounts seem to adhere to some sort of hierarchical response coding view such as the one proposed by Heister et al. (1990). According to Heister et al., responses are first and foremost coded and accessed by spatial (key) location codes. Whenever the spatial dimension allows discrimination of the responses, effector coding is assumed either (a) to simply contribute to a lesser degree (e.g., Hommel, 1993a; Worringham & Kerr, 2000), or (b) only to be included in a later stage of response selection or execution (i.e., when the response location codes are 'translated' into specific motor programs, leading to some sort of response-response (R-R) (in)compatibility; cf. Roswarski & Proctor, 2003a).

Recently however, Wascher et al. (2001) noted that position coding of responses is usually suggested by <u>instructions</u> and task demands in general. Specifically, standard instructions emphasize key positions by instructing subjects to press a particular (left or right) key in response to specific stimuli. Moreover, the stimulus-key mapping usually remains constant when hand condition is varied within subjects, whereas the stimulus-hand relations change. As would be assumed by the strong version of the direct coding hypothesis according to which response codes can be weighed as instructed, Wascher et al. argued that finger instructions (e.g., "respond with the left finger to the letter A"), combined with constant stimulus-to-finger mapping across hand conditions (implying varied stimulus-to-key mapping), should encourage subjects to more strongly rely on anatomical finger coding. Hence, a reduced or even reversed position-based effect should be observed in the crossed-hands condition under hand instructions. In contrast, such an outcome would not be expected by the spatial coding hypothesis, according to which the spatial (position-based) Simon effect under crossed-hands conditions should be largely unaffected by instructions.

Recent studies by Wascher et al. (2001) as well as Roswarski and Proctor (2003a) that compared hand and key instructions for crossed and uncrossed hands conditions speak to this issue. When using visual stimuli, Wascher et al. (Exp. 1) found no interaction between instructions, hand condition, and position compatibility. That is, the location-based Simon effect for crossed hands were 32 and 24 ms under key and hand instructions, respectively, and no overall hand-based Simon effect showed (i.e., conditions +/+ and -/+ as compared to conditions -/- and +/-, see Table 2). However, with auditory stimuli (Exp. 2), the expected outcome was obtained. That is, the position-based Simon effect in the crossed-hands condition was extremely reduced (3 ms) under finger instructions as compared to key instructions (36 ms), and an overall hand-based Simon effect of 19 ms was obtained. This finding seems at least partially consistent with the intentional weighing hypothesis in that finger instructions in the Wascher et al. experiment were able to reduce (but not reverse) the impact of position compatibility for auditory stimuli.

However, Roswarski and Proctor (2003a) were not able to replicate the Wascher et al. results. More specifically, they did replicate the lack of instructional influence with visual stimuli (Exp. 2) also observed by Wascher and colleagues (Exp. 1), but did not find a reduction of the position-based Simon effect for the crossed-hand condition under finger instructions with auditory stimuli (at least not in RTs, see their Exp. 3). Only when they substantially increased the number of trials and included session (practice) as a factor in the analysis of the results of their Experiment 4 (using finger instructions only), they observed a reduction of the position-based Simon effect with crossed hands in the second as compared to the first session (27 ms vs. 49 ms, respectively), whereas the Simon effect remained constant across sessions for the uncrossed hand condition. Roswarski and Proctor (2003a) concluded that spatial (location) coding is the default regardless of instruction, but that extensive practice in their Experiment 4 as well as in the Wascher et al. study, combined with finger instructions, may have encouraged stronger reliance on finger coding after practice. It remains unclear though, why this ‘practice effect’ was restricted to the finger instruction group, and why it did not occur with visual stimuli in the Wascher et al. (2001) study.

In sum, the results on finger vs. key instruction manipulations on the Simon effect with crossed hands seem to be more compatible with the spatial coding hypothesis than with the strong version of the direct coding (i.e., intentional weighing) hypothesis. First, with relatively limited practice (as in common behavioral experiments), the Simon-effect tends to be limited to stimulus-response location correspondence, and is little affected by type of response instructions and hand position. Second, for visual stimuli, even extended practice did not modify the location-based Simon effect for the finger-instruction group in the Wascher et al. study. Thus, subjects seem to prefer location-based (left-right) response coding, regardless of how responses are referred to in the response instructions.

As noted before, it remains unclear, why finger instructions modify the auditory Simon effect, and why such an ‘instruction effect’ only shows after extended practice. Moreover, the practice effect points to a more general problem inherent in all studies on the impact of instructions discussed in Chapter 3. This problem concerns the confounding of possible effects of instruction on (initial) response coding and changes in coding with practice that may or may not depend on instructions. In Chapter 5, I will present experiments that try to avoid this kind of confound.

While the findings presented in this section seem to favor the spatial (location) coding view when different spatial attributes of responses are emphasized by instruction, in the next section I will present evidence supporting arbitrary (non-spatial) response coding, both instructed and uninstructed.

3.1.3 Anticipated Action Effects

The evidence reviewed in this section is based on findings of so-called response-effect compatibility. 'Response effect compatibility' refers to the observation that stimuli presented after responding can come to influence responding. For example, Kunde (2001) had subjects respond to the color of centrally presented stimuli by pressing horizontally arranged keys. Pressing a key "produced" visual effects at different horizontal positions. In some conditions, the spatial relation between the manual response (i.e., the location of the finger and the key) and the visual effect was compatible, that is, the relative spatial locations of the responses and the response effects corresponded. In other conditions, however, responses and effects were incompatible. Kunde found slowed responses in the incompatible as compared to the compatible condition, even though the visual effect was presented after responding and was irrelevant for the task, indicating that anticipated action effects primed the (non-) corresponding response.

The question of interest here is whether response effect compatibility extends to more arbitrary (non-spatial) response-effects. If so, this would provide evidence for non-spatial response coding (see below).

Studies addressing this issue typically employ a two-step procedure including an acquisition phase and a test phase. In the acquisition phase, choice-responses (e.g., left and right keypresses) to arbitrary stimulus attributes are paired with some novel, arbitrary response-effect stimuli (e.g., color patches, tones of a certain pitch) that consistently succeed the responses (e.g., a left response is always followed by a red color patch, whereas a right response is always followed by a green color patch). In the subsequent test-phase the same responses (e.g., left and right keypresses) are required to imperative stimuli that often differ from those used in the acquisition phase. Importantly, stimuli resembling those that (formerly) served as response-effects are now used as primes (distractors) that are presented together with the imperative stimuli. The primes either correspond or do not correspond with the previously acquired effect associated with that response. If response effects were integrated into the response representation and used to access responses in the test phase, the primes should speed up or slow down responding. This is what ought to be expected according to both versions of the direct coding hypothesis, which assume that responses can be coded in terms of non-spatial features, in addition to spatial features. Moreover, the intentional weighing hypothesis predicts that the acquired action effect codes can be weighed more strongly than spatial response codes if responses are instructed in terms of their effects. Therefore, stimulus-response-effect compatibility can even be expected to override spatial correspondence effects.

In contrast, if responses are spatially coded regardless of response-effects and instructions, the arbitrary prime-distractors that correspond to the response effects should not affect responding.

Results obtained in studies on arbitrary (acquired) response-effect compatibility seem generally more consistent with the direct coding hypothesis in that they demonstrate arbitrary response priming. For example, Beckers, De Houwer, and Eelen (2002) demonstrated action-effect learning regarding affective evaluation of electrocutaneous feedback. In a training phase, participants moved a response key up or down in response to a go-signal. One of the responses was consistently followed by aversive electrocutaneous stimulation. In the test phase, word stimuli with a positive or negative connotation had to be classified according to their grammatical category (noun or adjective) by using the same responses (and response effects) as in the practice phase. Beckers et al. found a substantial valence-based compatibility effect (also termed ‘affective Simon effect’, cf. De Houwer & Eelen, 1998) even when the category-to-response mapping was regularly switched during the test phase, indicating that response valence had become integrated into the action representation and was used to control responding.

Similarly, Hommel (submitted) showed that subjects can access manual left and right responses via color codes. In his Experiment 1, participants first had to press a left or right key in response to letter identity of letters presented in a gray frame. Upon responding, the letter stimuli turned either green or red, depending on which key had been pressed. In a later test phase, red or green frames surrounding the letter stimuli were used as distractor stimuli.

Hommel found that responses were faster when the color of the distractor frame and the (learned) action effects (letters turning green or red) corresponded than when they differed, implying that the color effects had become integrated into the action representation and were used to access and guide manual responses. Similar effects have also been observed with other arbitrary effect-stimulus features such as pitch of tone (Elsner & Hommel, 2001; Hommel, 1996b), letter identity (Ziessler & Nattkemper, 2002), or the semantic category membership of words (Hommel, Alonso, & Fuentes, in press).

Whereas the findings reported thus far demonstrate that irrelevant action effects come to affect behavior, Experiments 2 and 3 by Hommel (submitted) suggest that the actual use of color codes may at least in part depend on their usefulness in a particular task, and hence on intentional weighing according to task demands. Hommel combined the manual color Stroop task (i.e., requiring left and right keypresses to the color of neutral words, or of color words written in a congruent or incongruent color) with color-related action effects (i.e., color frames or color words presented upon responding). He found a reduction of the manual Stroop effect (i.e., differences in performance as a function of word-color congruency) in RTs (but not in errors) for the group trained with a compatible color-effects mapping as compared to a control group without any response effects, but no statistically reliable difference between the control group and the incompatible color-effect group. Moreover, he observed a reduced (nonsignificant) impact of color-word action effects on the Stroop effect, leading him to conclude that the use of color-related action effects in response coding is under strategic control.

Taken together, these findings suggest that, in principle, abstract, non-spatial arbitrary codes can be used to control responding, implying considerable flexibility in response coding. This is consistent with the direct coding hypothesis by suggesting that other than spatial codes can be used to code and access responses. However, with the possible exception of Hommel (submitted) who provided evidence that the actual use of such effect codes may at least partially depend on the usefulness in a particular task, and hence on intentional weighing according to task demands, none of the studies reported so far speaks to the issue of whether task demands affect response coding. More specifically, none of these studies provides evidence for a direct influence of response instructions on response coding. Instead, they demonstrate non-instructed response coding, that is, the use of codes/dimensions that may have been primed through extensive practice and/or which may have proven useful for the task at hand.

The only study I am aware of that directly manipulated response instructions with respect to action effects of laterally organized manual responses has been conducted by Hommel (1993a; but also see Wang, Proctor, & Pick, 2002, for comparable findings with wheel rotation responses in a replication and extension of Guiard, 1983). Hommel (1993a) had participants react to the pitch of tones that were randomly presented to the left or right ear by pressing either a left or a right hand key. Pressing a key flashed on a light on the opposite side of the response key (so for example a left keypress would switch on a right light and vice versa). One group of participants was instructed to press the left vs. right key in response to tone pitch. This group produced a response-location-based Simon effect, that is, responses were faster when the location of the key corresponded to the side on which the tone was presented than when stimulus and response location did not correspond.

A second group of participants was instructed in terms of the light to be switched on. For example, they were told to respond to a low tone by switching on the light on the left. Under the latter task instruction the Simon effect reversed. That is, responses were now faster when tone location and the location of the light (that had to be switched on by the contralateral hand) corresponded. This finding implies that the response descriptions given in task instructions determined response coding, such that left responses were coded as right and vice versa.

It should be noted however that the reversed Simon effect under light instructions (-30 ms) was numerically smaller than the Simon effect under key instructions (52 ms). Similarly, the Simon effect under key instructions with incompatible key-to-light mapping was numerically smaller than the Simon effect observed for a control group (77 ms) that worked under key instructions with a compatible key-to-light mapping (i.e., for them, a left keypress switched on a left light and vice versa). These variations in the size of the effects are qualitatively similar to those observed in crossed-hands studies (cf. Chapter 3.1.2, above), presumably implying that the instructed codes were weighed more strongly than uninstructed codes, but that the latter may still have been part of the action representation, and hence contributed to the size of the (reversed) Simon effect by adding and/or subtracting from the overall effect.

To summarize, research on response-effect compatibility supports two important conclusions regarding the question of how responses are or can be coded. First, consistent with the direct coding view, studies investigating the acquisition of arbitrary irrelevant action effects show that response effects such as color become integrated into action representations and are or can be used to code and access responses in subsequent task performance. Second, the Hommel (1993a) results strongly suggest that instructions determine how responses are coded when salient coding alternatives are introduced by presenting clearly visible action feedback. When participants were instructed in terms of the locations of the lights turned on by a contralateral keypress, they relied more heavily on light location than key location (and anatomical hand) codes, whereas the key instruction group weighed response-location codes more strongly than light location codes.

This latter finding thus provides strong evidence in favor of the strong version of the direct coding hypothesis, which assumes that response codes can be weighed according to instructions. However, instructions in the Hommel study (as well as in the Wang et al. study) again emphasized different spatial aspects of the response array. Hence, one could argue that response instructions might have led to attentional shifts between different spatial aspects of the spatial dimension, thus affecting the hierarchy of spatial coding, but not the “prevalence” of spatial coding per se. More specifically, introducing salient lateralized action effects may have added another spatial dimension to the spatial coding hierarchy proposed by Heister et al. (1990), with key-location and effect-location being almost equally strong coding alternatives.

In the next section, I will review findings obtained with non-spatial response instructions. The majority of the studies used color instructions of responses, and originates from work on the Simon and the (manual) Stroop effect.

3.1.4 Non-Spatial Response Instructions

Compatibility effects with non-spatially instructed keypress responses have been studied with several paradigms, the majority of which instructed the buttons in terms of colors.

What would generally be expected according to the direct coding hypothesis is that compatibility effects should be observed with respect to the instructed dimension even when the overlapping stimulus attribute is task irrelevant. The strong version of the direct coding hypothesis furthermore predicts that instructed codes are weighed more strongly than uninstructed codes. Accordingly, non-spatial response instructions should reduce a given spatial compatibility effect.

In contrast, the spatial coding hypothesis predicts instruction independent spatial compatibility effects and a lack of irrelevant effects for instructed (non-spatial) response dimensions.

The paradigms involving non-spatial (color) instructions of keypress responses to be reviewed in this section include (a) the manual Stroop task, (b) the Hedge and Marsh task, and (c) color-compatibility effects with stimulus color as the irrelevant stimulus dimension. The results obtained with these tasks will be discussed in turn.

The Manual Stroop Task

The overwhelming majority of studies investigating the Stroop effect (e.g., Stroop, 1935) used verbal responses. In a typical verbal Stroop task, participants are required to name the color of incongruent color words (e.g., the word GREEN printed in *red*), the color of a control string (e.g., XXXX, or the word 'TABLE' printed in *red*), or the color of congruent color words (e.g., the word RED in *red* ink). The common finding with this task, the Stroop effect, is that it is much harder to name the color when the color is accompanied by an incongruent color word (e.g., GREEN printed in *red*) than to name the color of a colored control string (e.g., XXXX printed in *red*) or the color of a color-congruent word (e.g., RED printed in *red*).

This finding (for a comprehensive review, see MacLeod, 1991) has been interpreted as indicating response competition resulting from 'direct' response activation from the irrelevant stimulus (i.e., the color word). More specifically, some researchers (e.g., Cohen et al., 1990; Virzi & Egeth, 1985) have suggested that the irrelevant stimulus attribute (i.e., a color name) leads to response competition on incongruent trials, whereas it primes the correct response in the congruent condition.

Interestingly, the Stroop effect is extremely reduced or even nonexistent when the task is changed to word reading (i.e., when the relevant attribute is the color word and the irrelevant attribute the color of the print), indicating that there is some difference in type (e.g., Barber & O'Leary, 1997; Phaf, Van der Heijden, & Hudson, 1990; Virzi & Egeth, 1985) and/or strength (e.g., Cohen et al., 1990; Lu, 1997) of the pathways used to perform the word reading viz. color naming task (see below). That is, word reading is easier than color naming and less affected by irrelevant colors than vice versa either because of unequal (lifelong) practice, and/or because words and naming responses are of a similar 'format' and hence processed in the same system or module.

In the manual version of the Stroop task, verbal responses are replaced with keypress responses. Responses are instructed in terms of color, and the keys are either labeled with color patches or with color-words printed in black, allowing the comparison of the subtasks presented in Table 3, which result as a function of the relevant stimulus dimension (i.e., responding to color / ignoring color-names vs. responding to words / ignoring colors) and the type of the labels on the buttons (i.e., color patches vs. word-labels).

Table 3. Subtasks realized by different combinations of relevant (S_r) / irrelevant (S_i) stimulus attributes and key-label format in the manual Stroop task.

S_r	S_i	Label type	Response type	Translation
Color	Word	Color word	Translated word response	+
Color	Word	Color patch	Untranslated color response	-
Word	Color	Color patch	Translated color response	+
Word	Color	Color word	Untranslated word response	-

Note: 'Translation' refers to differences in format between relevant stimulus dimension and response label (terminology adopted from Sugg & McDonald, 1994).

To the extent that Stroop interference (only) measures response competition (e.g., Cohen et al., 1990) the spatial coding hypothesis predicts no manual Stroop effect. This is so because neither color words nor color patches should directly activate spatially coded responses. According to the direct coding hypothesis, on the other hand, a manual Stroop effect should be observed. Whether or not asymmetric effects should be expected for the translated versions of the two tasks (i.e., those for which the format of the relevant stimulus attribute and the key labels differ; see Table 3) and their untranslated counterparts depends on the assumptions regarding the role of perceptual and/or structural similarity. Although many models are silent or vague with respect to this question (but see Hommel, submitted) they nevertheless seem consistent with the view that perceptual and/or structural overlap leads to higher overall S-R overlap between stimuli and responses than conceptual overlap alone. If perceptual/structural overlap enhanced overall similarity, one would expect larger manual Stroop effects in translated as opposed to untranslated tasks, that is, similar asymmetries as the one observed in the verbal version of the Stroop task. Note that the strong and the weak version of the direct coding hypothesis seem to make the same predictions regarding this task because instructions refer to the key labels that are visible throughout the experiment.

The pattern of results typically obtained with manual responses can be described as follows. First, Stroop interference shows with both, verbal and color labels on the keys, although manual Stroop effects are typically smaller than verbal effects (cf. Sharma & McKenna, 1998;

see Sugg & McDonald, 1994, for an overview of findings obtained with different versions of the manual Stroop task). Importantly, Stroop interference tends to be modulated by task requirements, that is, by the combination of relevant stimulus dimension and label format (see Table 3). More specifically, pronounced effects are observed for both, the translated word-response-task in which subjects are required to respond to stimulus color (and to ignore the color-word) by pressing keys labeled with color-words, and the translated color-response-task that requires responding to color-words (and ignoring the color of the print) by pressing keys with color-patch labels. In contrast, usually no interference is obtained in the untranslated word-response task that requires word-label responses to the color-words irrespective of the color of the print (Pritchatt, 1968; Sugg & McDonald, 1994).

So far, the results of the manual task are consistent with and extend (by adding a translated color response task) those obtained with the verbal task, and seem generally consistent with the direct coding hypothesis. That is, under the assumption that key responses labeled with color <u>words</u> are primarily coded in terms of color <u>names,</u> direct activation from color-word distractors to responses (e.g., RED → “red”) is stronger than direct activation from color to color names (e.g., *red* → “red”), leading to more interference if the strong associate serves as distractor (i.e., in the translated word response task).

Similarly, the effect obtained in the translated color response task can be explained by assuming that key responses are primarily coded in terms of color or color concepts when keys are labeled with color patches. Interestingly however, most often significant (albeit typically smaller) Stroop interference is also observed for the untranslated color-response task that requires responding to color by pressing keys labeled by color patches (e.g., Keele, 1972; Pritchatt, 1968; Redding & Gerjets, 1977; Simon & Sudalaimuthu, 1979; Sugg & McDonald, 1994; White, 1969; but see McClain, 1983). In my view, the most plausible explanation for this outcome is that (automatic) lexical/verbal→semantic/conceptual activation (e.g., “red” → red) is stronger than semantic/conceptual→ lexical/verbal code activation (e.g., *red* → red → “red”; cf. Sugg & McDonald, 1994, discussion of Experiment 1). According to this account, color names automatically activate their corresponding concepts in the untranslated color-response task, leading to interference with the correct response that is also accessed via conceptual (and/or perceptual) codes. In contrast, lexical activation from conceptual codes representing irrelevant colors in the untranslated word-response task would be weak or nonexistent, leading to less interference[4].

In sum, if one assumes that Stroop interference is primarily due to response competition, as many researchers do (e.g., Cohen et al., 1990; Virzi & Egeth, 1985), then keypress responses must have been coded in terms of color, color names, and/or conceptual color codes. Color coding is induced by instructing participants to press a particular key that is labeled in a certain way. Hence, according to response-competition accounts of the manual Stroop effect, instructions (and key labels) influence response coding, thus supporting the direct coding view. Effects of label format are consistent with this view by demonstrating considerable flexibility of coding.

However, not all researchers agree with the response competition account of the Stroop effect. On the one hand, some researchers attribute the (manual) Stroop effect primarily to conceptual stimulus identification (e.g., Hasbroucq & Guiard, 1991; Kornblum & Lee 1995; Kornblum et al., 1999; Lu, 1997; Lu & Proctor, 2001), that is, to congruency viz. incongruency between the two stimulus attributes. According to this view, both verbal and color stimuli activate their corresponding concepts, thus hindering identification of the relevant stimulus or stimulus attribute in case of incongruent stimuli. Whereas 'pure' identification explanations do not seem to be particularly well suited to account for labeling-effects, translation models are.

Translation models (e.g., Glaser & Glaser, 1989; also see Sugg & McDonald, 1994, for an adapted version of the Glaser & Glaser model) emphasize the structural relationship between the relevant stimulus type and the response type. In general, they propose that colors and words are processed in different processing modules, each of which has its own codes. So, for example, according to Glaser and Glaser (1989; also see Phaf et al., 1990), color stimuli have privileged access to semantic (conceptual) codes, whereas word stimuli predominantly activate lexical (verbal) codes. According to translation models, substantial interference occurs if (a) information has to be translated to a code in the other system, and (b) irrelevant information has privileged access to the required code. Importantly, the primary source of interference according to these models appears to be activation (competition) at some in-

[4] A prominent alternative interpretation for untranslated color-effects states that stimuli and/or responses in the untranslated color-response task are re-coded into lexical representations, essentially transforming the untranslated color-response task into a doubly-translated word-response task (see, e.g., Hommel, submitted; Sugg & McDonald, 1994, for discussions). However, this explanation neither seems particularly plausible nor does it receive consistent empirical support.

termediate (lexical or semantic) stage, not necessarily response activation per se (cf. Sugg & McDonald, 1994). For instance, left responses to red stimuli in the translated word-response task (i.e., *red*→ ”red”→ left) are slowed because the distractor word GREEN activates a competing lexical representation (i.e., GREEN → “green” → right).

Thus, stimulus identification and translation accounts seem to share the view that the primary ‘locus’ of interference concerns some intermediate stage between perceptual stimulus identification and response selection, rather than response coding and response selection itself. Deciding between response competition and stimulus identification/translation accounts of the manual Stroop effect is difficult or even impossible because there is overlap between (a) the relevant and the irrelevant stimulus dimension, (b) the relevant stimulus dimension and responses, and (c) the irrelevant stimulus dimension and responses, leading to multiple possible sources of interference.

Moreover, response-coding alternatives are not assessed in the manual Stroop task. More specifically, it could be the case that responses are coded in terms of left and right (at least in situations where only two response alternatives and horizontal key arrangements have been used), independently of instructions and key labels. The manual Stroop task does not provide a means to rule out this possibility because spatial response coding, and its susceptibility to instruction, are not assessed in this task (but see Lu and Proctor, 1995, for a review of findings obtained with variants of the spatial Stroop task that requires naming or keypress responses to positions or position words).

In sum, multiple possible sources of interference as well as the lack of measures for spatial response coding defy a strong interpretation of Stroop congruency effects in favor of the direct coding hypothesis. Rather, the results obtained with the manual Stroop task seem to be somewhat uninformative with respect to the question of whether responses are arbitrarily coded when so instructed. Some of the criticisms regarding the Stroop task have been met by studies using the Hedge and Marsh task, a task named after the researchers that developed it (Hedge & Marsh, 1975). Findings obtained with this task are reviewed in some detail in the next section because they have been interpreted as major evidence for obligatory spatial response coding (e.g., Lu & Proctor, 1995).

The Hedge and Marsh Task

In the Hedge and Marsh (H&M) task, responses are instructed in terms of color, and spatial response coding is assessed by randomly varying stimulus position. More specifically, participants are required to respond to colored (e.g., *red* and *green*) stimuli that randomly appear to the left or the right by pressing lateralized response keys that are labeled with corresponding color patches (i.e., red vs. green). Hence, relevant and irrelevant stimulus dimension do not overlap in this task (but see Hasbroucq & Guiard, 1991, for a different view).

Typically, two different color-mapping conditions are compared. In the direct mapping condition, participants are either instructed to press, for example, the red key in response to the red stimulus (i.e., instructions do not explicitly mention the correspondence relationship; Hedge & Marsh, 1975; Proctor & Lu, 1999), or to press the key of the corresponding color (e.g., Hasbroucq & Guiard, 1991; Lu & Proctor, 1994; Proctor & Pick, 2003). In contrast, the reversed mapping condition requires opposite color responses (e.g., green key to red stimulus). Table 4 illustrates the resulting conditions in terms of color compatibility and (task-irrelevant) spatial compatibility as a function of position compatibility and mapping.

Table 4. Stimulus-response compatibility conditions realized by spatial compatibility under the direct (same color) and reversed (alternate color) mapping in the Hedge and Marsh (1975) task.

Mapping condition	Position compatibility	
	compatible	incompatible
direct	+/+	+/-
reversed	-/+	-/-

Note. + and - indicate compatibility and incompatibility between stimulus color and response color / stimulus position and response location, respectively.

The spatial coding hypothesis predicts a spatial compatibility effect (Simon effect) of typical size under both the direct and the reversed color mapping because irrelevant stimulus position should directly activate the correspondingly coded responses, regardless of instructions. Moreover, a color compatibility effect (i.e., faster responses under the direct than under the reversed color mapping) is also expected because a compatible color mapping should speed up translation in the conditional route.

The weak version of the direct coding hypothesis makes essentially the same predictions as the spatial coding hypothesis, albeit for different reasons. That is, if one assumes that responses in the H&M task are coded in terms of both, color and location (e.g., Zhang et al., 1999) both stimulus color and position should activate their corresponding response codes via the direct route, augmented by ‘controlled’ activation via the conditional route. Because the weak version of the direct coding hypothesis does not predict differential weighing of response codes, and because the models do not provide a means to predict the relative contribution of the direct and indirect routes to response activation in a principled way, the H&M task therefore cannot differentiate between the spatial coding hypothesis and the weak direct coding hypothesis.

The strong version of the direct coding hypothesis also predicts a color compatibility effect. In addition, if response codes can be weighed according to instructions no or extremely reduced spatial effects should be observed under both mappings.

The pattern of results usually obtained with the H&M task can be described as follows (see Figure 2, for a summary). First, responses are much faster in the two color-compatible conditions (i.e., under the direct mapping) than in the color-incompatible conditions (i.e., under the reversed mapping). The size of the color-compatibility effects in different studies (using visual stimuli) is typically much larger than the overall Simon effect and the Simon effect under the direct mapping.

Second, and more importantly, a Simon effect of normal size (i.e., about 25 ms) is typically observed under the direct mapping (involving a +/+ vs. +/- comparison, see Table 4). That is, responses are faster when stimulus position and response location correspond, even though stimulus position is task irrelevant and responses instructions do not refer to key location (e.g., De Jong et al., 1994; Hedge & Marsh, 1975; Lu & Proctor, 1994; Simon, Sly, & Vilapakkam, 1981). Because the response keys in the H&M task are not instructed with reference to their location, Lu and Proctor (1995) summarized that "spatial coding of the responses influences RT even when the response alternatives are not directly defined by spatial features", and concluded, "the Simon effect occurs whenever the response must be coded spatially." (p. 181)

Thus, the results obtained under the direct mapping apparently support the spatial and/or the direct coding hypothesis, and seem to provide evidence against intentional weighing.

However, the picture gets more complicated if one considers the two mapping conditions in combination and the reversed mapping condition in isolation. More specifically, when both mapping conditions are considered together, the usual finding is that the two compatibility effects interact. That is, whereas responses are faster when stimulus position and response location correspond (i.e., when they are spatially compatible) with the direct mapping, under the reversed mapping spatially compatible responses are <u>slower</u> than incompatible responses (see Figure 2). This pattern of results is difficult to reconcile with all three response-coding hypotheses that either predicted Simon effects in the usual direction or no spatial effect at all.

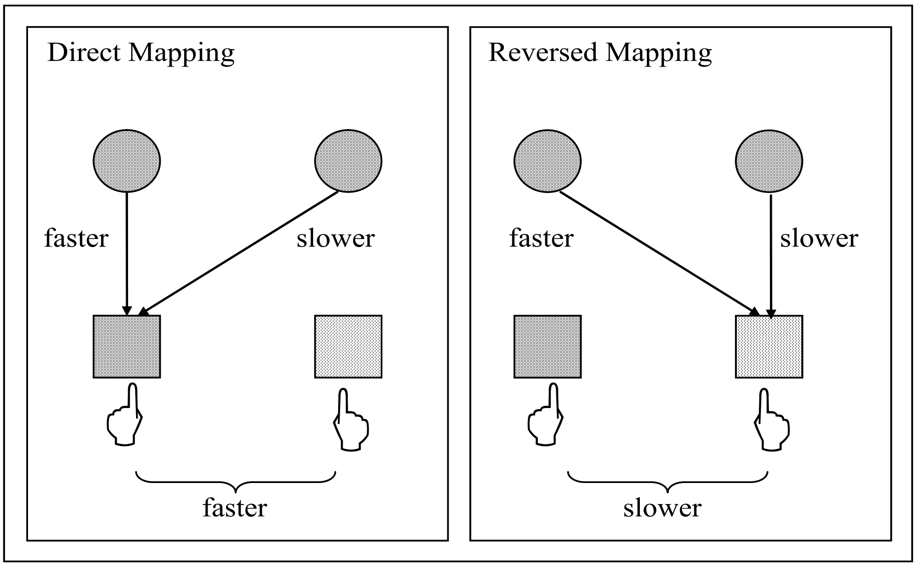

Figure 2. Schematic diagram of the basic findings in the Hedge and Marsh (1975) task under the direct (left panel) and the reversed (right panel) mapping (adopted from Lu & Proctor, 1995). *Note*. Different fillings indicate different colors.

Nevertheless, this finding is quite robust. It has been replicated in several studies (e.g., De Jong et al., 1994; Lu & Proctor, 1994), both with the original color-to-color mapping conditions, and with color-word response labels (i.e., with a variant of the translated word-response task, see previous section). Interestingly, Simon and Sudalaimuthu (1979) observed a similar RT pattern across mapping conditions with both an untranslated and a translated color-response version of the manual Stroop task with two response alternatives (i.e., faster responses with congruent distractors under the direct mapping as opposed to faster responses with incongruent distractors under the reversed mapping; see previous section for an overview of the different versions of the Stroop task), presumably implying that the H&M and the manual Stroop task share some properties.

The mechanisms underlying the impact of mapping instructions on the direction of the Simon effect in general, and especially its reversal under the reversed mapping, have been subject to considerable debate.

Two classes of explanations tend to be most prevalent to date, namely logical recoding accounts and an explanation in terms of display-control-arrangement correspondence (DCC)[5].

The basic tenet of the – arguably most widely accepted - logical recoding account (in the version of De Jong et al., 1994, and Lu & Proctor, 1994) is that participants recode the

[5] The stimulus identification account proposed by Hasbroucq and Guiard (1991) is omitted here because it seems as if it has been successfully rejected, both on theoretical and empirical grounds (for a comprehensive discussion, see Lu & Proctor, 1995). Moreover, it can be considered similar to the DCC hypothesis in some ways.

relevant and – inadvertently – also the irrelevant (spatial) stimulus attribute in order to form the response code. Under the reversed mapping condition a "respond opposite" rule is formed that transforms the values on both stimulus dimensions into their opposite. Consequently, a left stimulus attribute now primes a right response, leading to faster responses in the incompatible than the compatible condition.

Hardly surprising, the assumptions regarding what leads to the normal Simon effect under the direct mapping are more heterogeneous. According to some researchers (e.g., Hedge & Marsh, 1975; Lu & Proctor, 1994), an analogous (to the reversed-mapping condition) recoding rule, in this case some sort of "same" operation, is applied to both stimulus dimensions, leading to faster responses when both color and position correspond with the required response. This assumption seems plausible, given that (a) most studies manipulated mapping within subjects, and (b) the mapping instructions of many studies emphasized the correspondence viz. noncorrespondence between relevant stimulus and response dimension (see above). On the other hand, some researchers (e.g., De Jong et al., 1994) assume that, in addition to identity transformations, direct activation of compatible responses is responsible for the Simon effect under the color-compatible mapping condition, but not for its reversal under the reversed mapping.

This view gains support from distribution analyses of the effects and from studies that tracked response activation via lateralized readiness potentials (i.e., an electrophysiological index taken to reflect motor preparation processes). These studies show early activation of the spatially compatible response regardless of mapping (e.g., De Jong et al., 1994; Valle-Inclán, 1996), suggesting that fast (but not slow) responses are influenced by direct activation, leading to an enhanced Simon effect for the fast responses under the direct mapping, and to a reduced reversed Simon effect for faster responses under the reversed mapping (e.g., De Jong et al., 1994; but see Zhang & Kornblum, 1997, for a critique of this interpretation).

In contrast, the DCC account initially proposed by Simon et al. (1981) attributes the results to display-control arrangement correspondence. Such correspondence exists when the location of the color stimulus corresponds to the location of the same-color response-label (control). DCC is invariant across mappings. Consequently, DCC under the reversed mapping is present when stimulus location and response location are incongruent. So, for example, a red stimulus that requires a green/right response can be spatially aligned with the red/left response label when presented to the left, but not when presented to the right, leading to faster responses on spatially (S-R) incompatible trials.

Evidence in support of this view comes from two task modifications of the original H&M task. In one (e.g., Proctor & Pick, 2003, Exp. 1 and 2; Simon et al., 1981, Exp. 3), relevant and irrelevant stimuli are presented in different modalities. For example, color-responses are required to centrally presented color-stimuli, whereas irrelevant location information comes from tone stimuli that are randomly presented to the left or the right ear. With this arrangement, usually no spatial effect (i.e., no Simon-reversal) shows under the reversed mapping. Second, Proctor and Pick (2003) noted that, in the majority of studies using the H&M task, color labels were clearly visible during target processing. That is, most often they are presented below the stimuli in the lower part of the screen. When Proctor and Pick (Exp. 2) used key-labels that were not visible during stimulus processing, the Simon-reversal under the reversed mapping was absent, presumably implying that participants (perceptually) aligned stimuli and responses in the studies under visible-label conditions (but see De Jong et al., 1994, who found a reversal in the translated word-response task; also see Zhang, 2000, for critical findings).

Taken together, the Simon effect under the direct mapping and the results from distribution analyses (as well as from electrophysiological recording) obtained with the H&M task seem generally more consistent with the view that responses are spatially coded, either with (as assumed with by the weak version of the direct coding hypothesis) or without (as assumed by the strict spatial response coding view) color <u>response</u> codes contributing to the color compatibility effect. That is, the results seem to provide evidence against the strong direct coding hypothesis.

However, the origin of the reversal of the Simon effect under the reversed color mapping, and the origin of the normal Simon effect under the direct mapping is still unclear. That is, (non-) correspondence instructions as well as clearly visible response labels may have induced a bias to either logically recode stimuli (and/or responses) in terms of same/opposite rules, or use a perceptual matching strategy as proposed by the DCC hypothesis. Moreover, because color-overlap only existed in the task-relevant dimension, the color effect cannot unambiguously be interpreted as indicating color-coding of responses.

A stronger test of color coding of responses as well as intentional weighing would be provided by experiments that (a) investigated the influence of task-<u>irrelevant</u> stimulus-color on responses that are instructed in terms of color while avoiding S-S correspondence and assessing spatial coding, and/or (b) studied instructional modulation of (irrelevant) spatial stimulus-response compatibility by instructing responses spatially vs. non-spatially. It seems as if relatively few studies chose either of these approaches. Those I am aware of are reviewed in the next section. Again, irrelevant color effects would be expected according to the direct coding hypothesis. Moreover, the intentional weighing hypothesis predicts reduced to non-existent spatial effects under color (non-spatial) instructions of responses. In contrast, according to the spatial coding hypothesis, irrelevant stimulus effects should be restricted to the spatial dimension, and spatial effects should be unaffected by response instructions.

Response Instructions and Spatial vs. Non-Spatial Irrelevant Stimulus Effects

The evidence regarding response coding stemming from experiments that either assessed the impact of irrelevant stimulus color or investigated the spatial Simon effect under non-spatial response instructions appears to be mixed.

For example, Simon, Acosta, Mewaldt, and Speidel (1976, Exp. 2) conducted an experiment that provides evidence in favor of spatial response coding despite non-spatial response instructions. Simon et al. (1976) instructed participants to press a key of a certain color in response to the pitch of tone. Whereas a fixed-label group worked under a constant color-to-key assignment (i.e., a constant assignment of color to key-location), for other groups, the color-to-key mapping varied from trial to trial. In the varied color-to-key conditions, key labels were either presented 1s before the imperative stimulus (pitch), simultaneous with the stimulus, or after stimulus onset. Simon et al. found a significant Simon effect (i.e., a compatibility effect between irrelevant tone location and response side) both, with fixed key labels and in the varied color-to-key conditions that presented the labels prior to or simultaneous with the imperative stimulus. Because responses were instructed in terms of color (i.e., non-spatially), the Simon effect observed under these conditions seems to support the spatial coding view, and hence, the strict spatial coding hypothesis and/or the weak version of the direct coding hypothesis according to which implicit and explicit S-R overlap lead to comparable direct route activation.

However, although responses must have been coded spatially to some extent (otherwise no Simon effect would have been observed), the question of whether response coding is modulated by instructions remains unanswered by the Simon et al. (1976) results because the experiment did not include a condition with spatial response instructions. That is, the experiment does not allow a comparison of the size of the Simon effect under spatial vs. non-spatial response instructions. Moreover, participants in the Simon et al. (1976) experiment fixated the response arrangement during auditory stimulus processing, possibly encouraging some subjects to spatially re-code their responses (and hence, to partially ignore instructions), thus leading to the Simon effect.

Some evidence in favor of the spatial re-coding interpretation stems from a comparison between the fixed-label group on the one hand, and the 1-sec prior and simultaneous label groups on the other hand. Whereas the Simon effect in the fixed-label group was about 60 ms (a standard effect size with lateralized auditory stimuli; cf. Proctor & Pick, 2003), the Simon effect in the two varied-mapping groups was (marginally, but see sample sizes) reduced to 36 ms. Interestingly, the Simon effect in the 1-sec prior group and the simultaneous-label group was comparable even though overall RT level in the latter group was much higher, indicating that overall RT differences were not responsible for the difference in effect sizes between the latter two groups and the fixed-label group. Therefore, it seems possible that the fixed-label group recoded the instructions in terms of response location during the course of the experiments, leading to a more pronounced Simon effect.

Stronger evidence for the prevalence of spatial response coding has been provided by Brebner (1979) who switched the relevant and irrelevant dimensions in the H&M task (see previous section). More specifically, Brebner used a paradigm in which stimuli and responses did or did not correspond in terms of both, location and color, but unlike the original H&M task, participants were required to respond with the key on the same side (direct mapping) or the opposite side (reversed mapping) of the stimulus regardless of stimulus (and key) color. With this task, Brebner neither found a color-compatibility effect under the direct location-mapping, nor a reversed color effect under the reversed location-mapping. This result suggests that the irrelevant color dimension can be ignored more easily than irrelevant location with respect to whatever transformations or processes are required to perform the task under the different mappings.

However, Brebner (1979) has been criticized by Kornblum et al. (1990, p. 267) on methodological grounds. Moreover, results of studies that also used two-dimensional color-space arrangements, but instead instructed S-R mappings cross-dimensionally, lead to different conclusions with respect to the impact of instructions on response coding.

For instance, Hasbroucq and Guiard (1991) included two cross-dimensional mappings in their Experiment 2. In the color-to-position mapping, subjects were required to press the left or right key (color-labels of keys varied randomly from trial to trial) in response to stimulus color, and to ignore (randomly varied) stimulus position (for example, the instruction was "respond to a red stimulus by pressing the left key", thereby making stimulus position task irrelevant). Similarly, in their position-to-color mapping, subjects were required to press a key of a specific color in response to the location of a stimulus, with stimulus color being task irrelevant (e.g., "respond green to a left stimulus"). Again, color-labels randomly changed from trial to trial. Hasbroucq and Guiard (1991) found comparable irrelevant stimulus color and stimulus position compatibility effects (both about 50 ms) in the position-to-color and the color-to-position mappings, respectively. Moreover, they did not find an effect of irrelevant overlap between the relevant stimulus dimension and the irrelevant response dimension, at least not for the position-to-color mapping (i.e., no effect of spatial correspondence between relevant stimulus position and – randomly varying – irrelevant response location). This finding indicates that the instructed response dimension determines whether a stimulus-response compatibility effect occurs, thus supporting the strong version of the direct coding hypothesis. That is, these results support the view that non-spatial response instructions reduce the weights of spatial response codes, and hence the effect of (irrelevant) spatial stimulus attributes on responding.

Similarly, Smith and Brebner (1983) also manipulated mapping across dimensions. In their experiment, the (irrelevant) color compatibility effect in the position-to-color condition was even larger than position compatibility effect in the color-to-position condition, and a color effect even showed in the color-to-position condition, again seemingly supporting the strong version of the direct coding hypothesis.

However, both the Hasbroucq and Guiard (1991) and the Smith and Brebner (1983) study found slowed color-responses as compared to location-responses, possibly indicating an additional translation step. Moreover, both studies have been (e.g., Hommel, 1995), or can be, criticized for methodological reasons as well. For example, Hommel (1995) noted that participants in Hasbroucq and Guiard's Experiment 2 had to work under several mapping requirements in succession (i.e., color-to-color, position-to-position, color-to-position, and position-to-color), possibly leading to carry-over effects or specific strategies induced by previous task requirements. Similarly, Smith and Brebner's participants were required to work under both of two variants of either the color-to-position or the position-to-color mapping in succession. For example, in one variant of the position-to-color mapping subjects had to press the key of the same color as the stimulus when the stimulus appeared on the left, but to press the alternate color button when the stimulus appeared on the right. In the other variant however, the same/alternate assignment was reversed for the same subjects (i.e., they were now required to respond with the opposite-color button to a left stimulus, and to make same-color responses to right stimuli). This design presumably also led to pronounced inter-task transfer effects.

In sum, the studies reported in this section are highly inconclusive with regard to the main question of interest in this thesis, namely the question of whether response instructions determine how responses are coded. On the one hand, experiments with cross-dimensional mappings show symmetric (irrelevant stimulus-) color and position effects that depend on response instructions, apparently supporting the intentional weighing hypothesis, that is, the strong version of the direct coding hypothesis. However, these studies can be criticized because they used within-subjects manipulations of instructions, and because slowed responses on color-response tasks question the comparability of the position-to-color and the color-to-position task (but see Simon et al., 1976, who did not find an effect of overall RT level on the Simon effect in the two varied-label groups).

On the other hand, both the findings by Brebner (1979) and by Simon et al. (1976) apparently support the view that responses are spatially coded, regardless of response instructions. However, the Brebner (1979) study has been criticized for its lack of appropriate control conditions (cf. Kornblum et al., 1990). Simon et al. (1976), on the other hand, did not directly compare spatial and non-spatial response instructions, and their results with fixed vs. varied label-to-key mapping indicate that non-spatial response instructions (combined with varied labels) reduced the Simon effect observed in the fixed-label group who possibly (re-)coded their responses spatially. Therefore, their results cannot conclusively rule out the intentional weighing hypothesis (i.e., the strong direct coding hypothesis) either.

In Chapter 5, I present experiments that directly compare the Simon effect under spatial vs. non-spatial response instructions in order to more fully assess whether non-spatial response instructions reduce the Simon effect with spatial response instructions.

3.1.5 Summary

Stimulus-response compatibility effects are commonly attributed to a stage in information processing known as the 'response selection stage'. More specifically, dual route models assume that they result from response competition induced by response activation via a conditional (controlled) route that depends on mapping instructions, and a direct route that solely depends on overlap between stimulus and response codes. Accordingly, investigating which match or compatibility relations contribute to compatibility effects under which response instruction conditions allows conclusions about the cognitive codes that are used to control instructed responding.

Therefore, Chapter 3.1 provided a review of findings regarding the impact of (response) instructions on several SRC effects involving spatially organized keypress responses. This review was guided by the question of whether the results obtained with several tasks provide evidence for or against the coding hypotheses identified in Chapter 2.

According to the direct coding hypothesis, it was expected that variations in response instructions affect response coding, and hence, how responses are or can be accessed. Whereas the weak version of the direct coding hypothesis does not discriminate between implicit and explicit (instructed) overlap, the strong version assumes that response codes can be weighed according to instructions. Consequently, both versions predict compatibility effects resulting from stimulus-overlap with instructed (including non-spatial) response dimensions even when the overlapping stimulus attribute is task irrelevant. However, only the strong version of the direct coding hypothesis assumes that instruction manipulations should lead to variations in the direction or size of a given spatial compatibility effect. In contrast, the spatial coding hypothesis predicted (irrelevant) spatial compatibility effects of comparable size and in the same direction regardless of the specific contents of the response instructions. Other than spatial compatibility effects should be restricted to the task-relevant dimension, that is, they should be attributable to translation efficiency in the conditional route.

Given these criteria to decide between the different views, the general conclusion I arrived at has been that the findings obtained with the tasks reviewed in Chapter 3.1 provide highly inconsistent and/or ambiguous evidence for and against all hypotheses.

Sections 3.1.1 – 3.1.2 were mainly concerned with instruction manipulations according to which different instruction conditions emphasized different spatial aspects of the stimulus- and/or response array.

With the two-dimensional spatial mapping task (see Chapter 3.1.1) a weak left-right prevalence effect is typically observed. That is, although horizontal and vertical compatibility effects tend to be modulated by type of instruction (i.e., vertical vs. horizontal instructions) in most studies, the overall horizontal compatibility effect is usually larger and less reduced by vertical instructions, indicating a dominance of left-right over top-bottom coding, and hence, apparently providing evidence in favor of the spatial coding hypothesis (in its one-dimensional interpretation).

However, it has been argued that, in this task, the two stimulus-dimensions (i.e., horizontal and vertical position) always redundantly signal the correct response, thus inviting different types of strategic recoding biases (i.e., re-interpretations of instructions). Moreover, because the relevant stimulus attribute in this task is always position, it cannot be determined whether this bias refers to stimulus and/or response coding, rendering this task less than optimal to answer the question of whether response instructions determine response coding.

The findings regarding the impact of response location vs. finger instructions in auditory and visual Simon tasks requiring crossed-hands responses (see Chapter 3.1.2) are inconclusive as well. Whereas Wascher et al. (2001) found an impact of response instructions on the pattern of hand-based vs. location-based Simon effects in the auditory task, Roswarski and Proctor (2003a) only found a small impact of anatomical coding under finger instructions after considerable practice. Moreover, instruction effects were negligible for visual tasks in both studies. Thus, although the results favor hierarchical spatial coding (with location-coding on the top of the hierarchy; e.g., Heister et al., 1990), and hence tend to speak against the intentional weighing (i.e., the strong direct coding) hypothesis, the reason for (a) the different pattern of results observed across stimulus modalities and (b) the effect of practice on response (re-?) coding in the auditory task remains unclear.

The strongest evidence in favor of intentional weighing of response codes (i.e., the strong version of the direct coding hypothesis) has been provided by Hommel (1993a) who instructed responses either in terms of response location or in terms of the location of contralaterally presented response effects (see Chapter 3.1.3). Hommel found that the response-location based Simon effect was partially reversed for the effect-instruction group (e.g., left keypress responses were now faster when the stimulus and the response effect appeared on the right, that is, when the stimulus appeared at a response-location incompatible position), suggesting that instructed response features were weighed more strongly than uninstructed response features. However, instructions in the Hommel (1993a) study emphasized different spatial aspects of the response array, thus possibly affecting the hierarchy of spatial coding, but not necessarily the prevalence of spatial coding per se.

Therefore, it has been argued that SRC effects resulting from non-spatial irrelevant stimulus overlap with the instructed response dimension would provide more stringent evidence for the direct coding hypotheses.

Evidence regarding non-spatial response coding as a consequence of non-spatial response instructions and labels (Chapter 3.1.4) seems ambiguous as well, though.

On the one hand, color-compatibility effects obtained with manual versions of the Stroop task and with the H&M task are consistent with the interpretation that responses were at least partially coded in terms of color (or color names).

On the other hand, in the manual Stroop task, spatial response coding has not been assessed, and researchers do not agree regarding the source(s) of interference in the Stroop task. More specifically, (conceptual) stimulus identification and/or interference at some intermediate translation stage cannot be ruled out as explanations because there is not only overlap between (irrelevant) stimulus attributes and responses, but also between the relevant and the irrelevant stimulus dimensions.

In contrast to the manual Stroop task, the H&M task avoids overlap between the relevant (color) and the irrelevant (position) stimulus attributes, and does provide a means to measure spatial response coding. However, the results obtained with this task are partially (in)consistent with all three hypotheses. First, the color compatibility effect does not differentiate between the alternative hypotheses because overlap exists regarding the task-relevant dimension. Second, whereas the (location-based) Simon effect typically observed under the direct mapping seems to indicate spatial response coding, and hence to support both the spatial and/or the weak direct coding hypothesis, the origin of the Simon effect in this task and its reversal under the reversed-color mapping are unclear. More specifically, the pattern of results is also consistent with an explanation in terms of a (strategic) bias to either perceptually align stimuli and response labels, as proposed by the DCC account, or to logically recode the mappings into "same" and "opposite" rules that are inadvertently applied to the irrelevant (spatial) dimension as well.

Studies that investigated irrelevant color effects with cross-dimensional instructions in the H&M task (Hasbroucq & Guiard, 1991; Smith & Brebner, 1983) seem generally better suited to assess response coding in terms of color. These studies found a symmetric influence of irrelevant stimulus position and irrelevant stimulus color that depended on response instructions. Whereas these results seem to favor the strong direct coding hypothesis, both studies using this type of mapping manipulation have been severely criticized on methodological grounds, again defying any firm conclusions.

Hence, it seems as if the strongest evidence in favor of arbitrary (i.e., non-spatial) response coding to date has been provided by experiments on response-effect compatibility (see Chapter 3.1.3). These studies demonstrate that arbitrary response effects such as color, pitch of tone, or affective valence, become integrated into the action representation, and are or can be used to code and access responses during subsequent task performance. However, none of these (arbitrary response-effect) studies provides evidence for or against a direct impact of response <u>instructions</u>. Rather, they demonstrate non-instructed response coding, that is, the use of codes that may have been primed by practice and/or that may have proven useful for the task at hand.

Finally, Simon et al.'s (1976, see Chapter 3.1.4) finding of a Simon effect under conditions where stimulus attributes (i.e., pitch of tone) were arbitrarily mapped to color-responses (i.e., when responses were non-spatially instructed) indicates that responses are at least partially coded in terms of location in their experiment, apparently contradicting the strong version of the direct coding hypothesis. However, their study did not include a condition with spatial response instructions. It seems possible that the Simon effect under spatial response instructions would be larger than under non-spatial response instructions. Therefore, intentional weighing of response codes according to instructions, and hence, a reduced impact of spatial response codes under non-spatial response instructions, cannot be precluded.

The experiments presented in the empirical part of the thesis (Chapters 4 and 5) attempt to assess more directly inhowfar the response labels used in verbal task instructions determine response coding. More specifically, they extend the evidence for arbitrary response coding obtained in studies on response-effect compatibility by addressing whether arbitrary (non-spatial) response coding occurs as a function of response instruction (Chapter 4), and they extend the Simon et al. (1976) results by directly comparing (irrelevant) spatial correspondence effects under spatial vs. non-spatial response instructions (Chapters 4 and 5). Moreover, in the experiments presented in Chapter 5, an attempt was made to avoid the confound between instructions and practice present in all prior studies so far (see Roswarski & Proctor, 2003a, for results indicating an impact of practice on coding) by introducing new imperative stimuli (and mapping instructions) on each trial.

Whereas the experiments reported in Chapter 5 rely on a Simon-type task and hence – methodologically speaking – directly relate to the findings reviewed in Chapter 3.1 above, the experiments presented in Chapter 4 used a dual-task paradigm that involved consistent viz. inconsistent responses on the two tasks (i.e., R-R compatibility).

Therefore, before proceeding to the empirical part of this thesis, Chapter 3.2 provides a review of findings concerning inter-task consistency effects similar to those presented in Chapter 4.

3.2 Response Instructions and Cross-Task Compatibility

'Inter-task consistency' or 'cross-task compatibility' (CTC) refer to the finding obtained in dual task studies, which require participants to simultaneously perform two tasks, that responding on either task is easier when the mapping for the other task requires consistent rather than inconsistent responses (see Lien & Proctor, 2002, for a recent review).

For example, Hommel (1998, Experiment 1) had participants perform a manual (primary) and a verbal (secondary) task in response to centrally presented visual stimuli. The (primary) manual task was to press a left or a right key in response to the color of colored letters, whereas the verbal (secondary) task required saying either "left" or "right," depending on letter identity. Hommel found R-R compatibility effects on both tasks, that is, both manual and verbal responses were faster when response 'locations' corresponded (e.g., faster responses when a left keypress was followed by a "left" as opposed to a "right" verbal response) than when they did not correspond.

This result has several important implications. First, response selection for the two tasks must have overlapped in time to produce backward (R2-R1) compatibility effects from Task 2 to Task 1. This finding contradicts strong response-selection-bottleneck accounts of the 'psychological refractory period' (PRP) effect, that is, the finding that R2 responses are particularly delayed when S1-S2 stimulus-onset asynchrony (SOA) is short. More specifically this finding is inconsistent with the proposal that a response selection bottleneck leads to complete postponement of S2-R2 translation until R1 selection has finished (see Lien & Proctor, 2002; Pashler, 1994, for comprehensive descriptions of the PRP effect and bottleneck as well as alternative accounts thereof). Rather, it indicates that responses on the two tasks were activated automatically and in parallel even though the S-R mappings for both tasks were arbitrary.

Second, the finding of inter-response effects between physically dissimilar responses such as pressing a key and saying a word indicates that participants used overlapping codes when accessing their verbal and manual responses. Under the assumption that cross-task compatibility in the Hommel experiment reflects parallel activation of responses (i.e., response codes) the results suggest that relatively abstract conceptual representations of space were used to code both verbal and manual responses (see General Discussion of Experiments 1-3, Chapter 4, for a more detailed discussion). Moreover, whereas forward R-R compatibility effects can be explained by some sort of (meaning based) automatic priming and thus might only indicate that the overlapping code had been part of the response representation of the secondary task, the finding of forward and <u>backward</u> compatibility effects suggests that these codes were not only part of both response representations, but that they were actively <u>used</u> to access and guide responding on the secondary task.

Inasmuch as such inter-task consistency effects reflect response-related processes, the nature of such effects and their susceptibility to response instructions again allow inferences about the codes involved in response activation and selection. If one generalizes the response coding hypotheses derived above to dual-task performance (i.e., by assuming some sort or R-R priming when response codes for the two tasks overlap)[6], the spatial coding hypothesis would again predict only spatial R-R compatibility effects that should occur independently of the response labels used in manual task instructions. In contrast, according to the direct coding hypothesis, one would expect cross-task compatibility for instructed (non-spatial) response dimensions as well. Moreover, the strong version of the direct coding hypothesis again predicts that cross-task compatibility should primarily be observed for the instructed response-dimension, regardless of whether it is spatial or not.

To anticipate the conclusions drawn in this section, the results on inter-task consistency effects so far seem to support the view that such effects depend on the instructed S-R <u>map-</u>

---

[6] Strictly speaking, R-R effects arising from overlap between responses on two concurrently performed, mostly arbitrarily mapped tasks, are not within the scope of explanation covered by current dual route models. However, they can easily be adapted by allowing (a) relatively fast development of direct S-R links between conceptually dissimilar stimuli and responses with practice (e.g., Hommel & Eglau, 2002; Proctor & Lu, 1999) leading to direct response activation whenever the response-associated stimulus is presented, and/or (b) strong automatic and parallel response activation via conditional routes (e.g., Lien & Proctor, 2002; Tagliabue et al., 2000; see Hommel, 2000, for a comprehensive discussion of different types of automaticity).

pings. However, they do not support a strong interpretation in terms of instructional impact on response coding.

For instance, findings by Lien and Proctor (2000) as well as Koch and Prinz (submitted) corroborate the conclusion that CTC effects as those found by Hommel (1998) are primarily based on parallel response code activation, and extend the Hommel results by showing that the direction of such effects depends on the instructed S-R mapping.

Lien and Proctor demonstrated that, at short SOAs, response selection on an arbitrarily mapped primary task is influenced by correct R2 activation even when R2 is mapped to S2 in a spatially incompatible way (e.g., when Task 2 required left responses to right arrows). For example, a left response to the letter “x” on the primary task was faster when it was accompanied by an arrow pointing to the right (requiring a left response on Task 2) than when it was followed by a left arrow, although the R2-R1 compatibility effects tended to be numerically smaller under the reversed than under the direct mapping (i.e., when responses on Task 2 were compatibly mapped to arrow direction). Furthermore, Lien and Proctor observed that R1 selection was affected by irrelevant arrow (S2) position in a similar way as R2 was. That is, there was a small backward “Simon” effect under the direct Task 2 (arrow) mapping, whereas the irrelevant location effect was slightly reversed under the reversed mapping. These results suggest that Task 2 responses instead of Task 2 stimuli primed Task 1 responses, and that it is primarily the instructed Task 2 mapping that contributes to inter-task consistency effects.

Similarly, Koch and Prinz (submitted, see also Koch & Prinz, 2002) who used a somewhat different methodology and varied the encoding instructions of an unspeeded perceptual identification task were able to show that the direction of CTC effects between the identification task and a nested, but logically independent choice reaction task depended on the encoding instructions of the perceptual identification task. In their study, one group of participants was instructed to report the starting point of the movement of a moving target stimulus, whereas another group was required to report the endpoint of the movement. Hence, the correct answer to a target moving to the left was “left” for the endpoint group, but “right” for the starting point group. Koch and Prinz found that choice task responses (speeded left/right finger movements to predetermined keys in response to the color of letter stimuli presented before the moving target) were faster when the direction of the finger movement corresponded to the movement aspects that had been emphasized by the instruction.

Thus, the findings obtained by Hommel (1998), Koch and Prinz (submitted), and Lien and Proctor (2000) indicate automatic response activation and inter-response priming that follows instructed S-R mappings. However, although these results show an impact of mapping instructions, they cannot conclusively differentiate between the different coding hypotheses. This is so because they did not directly manipulate response instructions. That is, responses were always instructed in terms of left and right, thus strongly encouraging spatial response coding. Accordingly, all three hypotheses, including the spatial coding hypothesis, are consistent with the results. Therefore, demonstrations of inter-task compatibility for other than spatial response dimensions, as well as studies that investigated whether consistency effects are restricted to the instructed response dimension appear to be better suited to address the question of the impact of instructions on response coding.

Logan and colleagues (Logan & Schulkind, 2000; see also Logan & Gordon, 2001) extended the findings regarding CTC effects reviewed so far by generalizing forward and backward consistency effects to non-spatial stimulus and/or response dimensions. Moreover, they were able to show that the occurrence of such inter-task consistency effects depends on the overlap of instructed categorizations.

For example, participants in Logan and Schulkind (2000, Experiment 2) had to categorize two numbers that were presented with varying SOAs. In some sessions, participants had to classify the numbers according to the same categories, that is, they either had to judge magnitude or parity on both tasks. In other sessions however, the categorization task varied from Task 1 to Task 2 (i.e., from magnitude judgments on Task 1 to parity judgments on Task 2 or vice versa).

Logan and Schulkind observed forward as well as backward category matching effects over a wide range of SOAs when the same categorization was required on the two tasks (e.g., responses were faster when both tasks required parity judgments and the numbers presented on Task 1 and Task 2 were both odd as opposed to one being odd and the other being even). Interestingly, however, no consistency effects (neither forward nor backward) were obtained when the two tasks required different stimulus categorizations (i.e., when one task required magnitude judgments while the other required parity judgments) although stimuli could still be classified according to both categories. So, for example, presenting the digit “4” (for Task 2) did interfere with Task 1 categorization of the digit “8” as being larger than “5” when task-

2 also required a magnitude judgment (i.e., a "smaller" response), but it did not interfere with Task 1 performance when it required a parity judgment (i.e., an "even" response).

Logan and Gordon (2001) interpret these and similar results as indicating that parallel activation or retrieval of response-relevant information depends on the amount of overlap of task-relevant response sets. One interpretation of this account (cf. Chapter 2.1) holds that responses are coded in terms of what they signal (e.g., a left response as meaning 'odd'; for a similar view, see e.g., Meiran, 2000). According to this view, the Logan and Schulkind results appear to support the strong version of the direct coding hypotheses.

However, in the experiments described by Logan and colleagues the two tasks were mapped onto different hands, and the contribution of responding at corresponding vs. noncorresponding relative response locations to the category matching effect (is it negligible, additive, or does it interact?) has not been assessed. Moreover, stimuli were bivalent in that each stimulus provided evidence for categorizations on both tasks (e.g., the stimulus '7' is both larger than 5 and odd). Hence, the contribution of stimulus-related processes, such as semantic priming (category-category priming) and <u>stimulus</u>-categorization processes to their findings is probably substantial. That is, it remains unclear to what degree, if at all, <u>response</u> representations were responsible for their results (but see Schuch & Koch, submitted; Watter & Logan, 2001, for promising attempts to disentangle stimulus- and response-related processes in the magnitude/parity judgment task).

Taken together, the findings reviewed in this section allow several conclusions with respect to response coding. First, backward compatibility of the kind observed by Hommel (1998), Lien and Proctor (2000), as well as Logan and colleagues, seems to be well suited to examine the actual <u>use</u> of specific response codes in Task 2 performance. In this regard, backward compatibility extends findings of forward compatibility (and, in a sense, also findings on stimulus-response compatibility, see Chapter 3.1) that may only show that the overlapping codes have been part of the Task 2 action representations.

Second, inter-task consistency effects indicate that conceptual – possibly arbitrary (e.g., Logan & Schulkind, 2000) – codes can be used to access physically dissimilar responses such as verbal and manual responses, or left-right responses with different hands.

Third, the results obtained by Koch and Prinz, Lien and Proctor, as well as by Logan and colleagues suggest that inter-task consistency depends on task demands, that is, on the instructed mapping rules. However, they do not support a strong interpretation in terms of an instructional impact on response coding because either (a) only spatial response instructions were used for both tasks (Koch & Prinz, submitted; Lien & Proctor, 2000), or (b) because spatial response coding has not been assessed (e.g., Logan & Schulkind, 2000). It has been argued that the latter does not allow firm conclusions in terms of arbitrary response coding, whereas the former does not conclusively differentiate between the alternative coding hypotheses.

A more straightforward way to address the question of whether response instructions directly influence response coding of spatially organized responses (i.e., left-right responses on a manual task), would be to use univalent stimuli and arbitrary S-R mappings (as, for example, in the Hommel, 1998, study) and to vary the response labels used in the instructions of a manual (secondary) task. On the one hand, such an approach allows less ambiguous conclusions as to whether inter-task consistency effects generalize to more abstract (non-spatial, e.g., color) response instructions, and hence, non-spatial response codes when a concurrently performed (e.g., verbal) task also requires this type of code (e.g., color codes). This would provide evidence in favor of direct coding in general.

On the other hand, such an approach also allows testing whether instructed response codes can override spatial response coding by requiring spatial coding on a primary (e.g., verbal) task and non-spatial coding of (spatially organized) responses on the secondary task. Whereas both the spatial and the weak direct coding hypothesis predict that the spatial backward-compatibility effects from a manual keypress task should be unaffected by response instructions, the strong version of the direct coding hypothesis predicts reduced spatial (forward and) backward effects under non-spatial manual response instructions.

This rationale is exactly the logic underlying the experiments in Chapter 4 that will be presented after summarizing the general aims of the study.

3.3 Summary and Aims of Study

Chapter 3 provided a review of findings that speak to the main question of this thesis, namely whether or not the specific response labels (e.g., “left” and “right”) given in simple binary choice task instructions involving spatially organized keypress responses (e.g., “when you see a square on the screen, press the left key; when you see a circle, press the right key”) determine how such a task is performed, that is how responses are coded and selected.

To this end, an overview of findings on the impact of response instructions on a number of different stimulus-response and inter-task compatibility phenomena involving spatially organized keypress responses has been provided. The main question throughout this literature review was whether variations of response instructions affect the type, size or direction of observed compatibility effects. More specifically, according to the direct coding hypothesis derived in Chapter 2, it was expected that variations in response instructions affect response coding, and hence, how responses are accessed. Whereas the weak version of the direct coding hypothesis does not discriminate between implicit and explicit (instructed) overlap, the strong version assumes that response codes can be weighed according to instructions. Consequently, both versions predict compatibility effects resulting from overlap on the instructed dimension even when the instructed response dimension is non-spatial and the overlapping stimulus or response attribute is task irrelevant. However, only according to the strong direct coding view should instruction manipulations lead to variations in the direction or size of a given spatial (implicit) compatibility effect.

In contrast, the spatial coding hypothesis predicted (irrelevant) spatial compatibility effects of comparable size and in the same direction regardless of the specific contents of the response instructions. If other than spatial compatibility effects occurred, they should be attributable to translation efficiency in the conditional route.

The main conclusion derived from both, the stimulus-response and the response-response compatibility literature has been that the results are largely ambiguous with respect to the question of how instructions influence response coding for several reasons.

First, findings obtained with a wide variety of paradigms suggest that instructed S-R <u>mapping</u> affects how a task is performed. For example, an impact of task demands on task performance has been observed with the Hedge and Marsh task (direct vs. reversed mapping), the two-dimensional mapping task (vertical vs. horizontal stimulus position to response location mapping), different versions of the manual Stroop task (translated vs. untranslated versions of the color-response and the word-response task), and in dual task studies with overlapping viz. non-overlapping responses on the two tasks (e.g., direct vs. reversed Task 2 mapping in the Lien & Proctor, 2002, experiments).

However, these findings cannot unambiguously be attributed to an effect of <u>response</u> instructions on <u>response</u> coding for several reasons. One reason is that, in most of these tasks, there was not only overlap between the irrelevant stimulus dimension and the instructed vs. uninstructed response dimensions, but there was also overlap between the relevant stimulus dimension and the (un)instructed response dimension (in the H&M task, the two-dimensional mapping task and the manual Stroop task), and, in some cases, also between the relevant and the irrelevant stimulus dimensions (in the manual Stroop task). Moreover, several of the tasks that indicate an impact of irrelevant non-spatial information when responses are instructed non-spatially (i.e., the manual Stroop task and the experiments by Logan and colleagues) did not assess spatial coding.

It has been argued that these tasks do not permit firm conclusions as to the sources (i.e., the locus) of interference. More specifically it cannot be precluded that non-spatial compatibility effects were due to some intermediate translation (or stimulus-recoding) stage that transforms stimuli onto left/right (i.e., spatially coded) responses. Therefore, they cannot firmly rule out the spatial coding hypothesis.

The dual task studies that showed inter-task consistency effects involving left-right responses to depend on instructed S-R mapping (e.g., Koch & Prinz, submitted; Lien & Proctor, 2000), on the other hand, are uninformative because they did not vary response instructions or labels (i.e., they only used spatial response instructions). Because all three hypotheses make the same predictions regarding spatial response instructions, these results do not directly speak to the issue of how response instructions affect <u>response</u> coding.

Second, those studies that varied response instructions directly and studied the impact of irrelevant stimulus attributes as a function of <u>response</u> instructions lead to inconsistent conclusions that depend on the task (and the instructions) used. Regarding instruction manipulations that emphasized different spatial aspects of the response array, results on anatomical vs. location instructions provide inconsistent results that seem at least slightly more consistent with the spatial coding hypothesis, that is, with the view that location-based coding dominates regardless of instructions. In contrast, instructions that either emphasized response location or salient contralateral response effects (e.g., Hommel, 1993a) support the strong version of the direct coding hypothesis by showing that instructed codes were weighed more strongly than spatial (key location) codes. Participants in the contralateral-effect instruction-group must have coded their left and right location responses as right and left, respectively, in order to produce a reversed Simon effect. While the Hommel (1993a) result is probably the strongest evidence in favor of intentional weighing of codes obtained so far, response instructions for both the response location and the response-effect group referred to the spatial (i.e., left/right) dimension.

However, studies that either investigated the impact of irrelevant stimulus location on non-spatially instructed responses for arbitrary S-R mappings (Simon et al., 1976), or that studied the impact of irrelevant stimulus information that overlapped with the <u>instructed</u> response dimension with cross-dimensional mappings in the two-dimensional color-location task (Hasbroucq & Guiard, 1991; Smith & Brebner, 1983) have been criticized for either a lack of appropriate control conditions (i.e., the lack of a condition with spatial response instructions in the Simon et al., 1976, experiment), or for methodological reasons (i.e., within-subjects manipulations of instructions).

Therefore, the results reviewed in Chapter 3 must be considered inconclusive with respect to direct viz. spatial response coding, at least where non-spatial response instructions are concerned.

The experiments presented in Chapters 4 and 5 extend the existing findings in several regards. They are similar to the Hommel (1993a) approach in varying response instructions on otherwise identical (or at least very similar) tasks. However, unlike Hommel, response instructions in my experiments did not emphasize different spatial aspects of the response array and did not present stimulus-compatible viz. –incompatible response-effect stimuli.

More specifically, the general logic underlying the experiments was to vary response instructions for manual (left and right) keypress responses to arbitrary stimulus attributes. This was done by instructing the response keys as either left vs. right keys (spatial instructions) or as blue vs. green keys (color instructions). If participants arbitrarily code and access their responses as instructed, response instructions should (co-) determine how responding is controlled.

By using color instructions and labels I rely on the findings provided by experiments on response-effect compatibility, which demonstrate that – in principle – manual keypress responses can be arbitrarily coded and assessed. However, unlike studies on response effect compatibility, participants in my experiments were not presented with (i.e., trained on) arbitrary color effects, but instead were provided with color-labels before each trial (Chapter 4) or before each block of trials (Chapter 5).

Two different experimental approaches were used to assess whether response instructions determine response coding. Both rely on the compatibility logic outlined in Chapters 2 and 3 and used arbitrary relevant S-R mappings in order to avoid simultaneous overlap on more than one dimension.

In one set of experiments (Experiments 1-3, Chapter 4), a dual task methodology similar to that used by Hommel (1998, Experiment 1) was employed. More specifically, in addition to a (secondary) manual keypress task with varied response instructions, participants had to perform a (primary) verbal task that either required “left” vs. “right” or “blue vs. green” concurrent verbalizations. Experiment 1 was a conceptual replication of the Hommel (1998) experiment, in which responses on both tasks were instructed in terms of left and right. This experiment served as a ‘baseline’ for Experiments 2 and 3. In Experiment 2, both verbal and manual responses were instructed in terms of color. If instructed response codes are or can be used in manual response coding, as assumed by both versions of the direct coding hypothesis, then (forward and) backward compatibility effects should generalize to the color dimension. Experiment 3 was designed to differentiate between the strong version of the direct coding hypothesis, on the one hand, and the spatial and weak direct coding hypothesis, on the other hand. This was done by again instructing manual responses in terms of color, but requiring location coding (i.e., “left” and “right” responses) on the verbal task. According to both the weak version of the direct coding hypothesis and the spatial coding hypothesis, non-spatial response instructions should not affect spatial coding of manual responses. Hence, spatial R-R compatibility effects similar to those in Experiment 1 should be observed. In contrast, the strong version of direct coding hypothesis predicts that instructed codes are weighed more strongly. Therefore, reduced spatial inter-task effects should be obtained.

By instructing the responses on both tasks differently, Experiment 3 can also be considered an extension of the Logan and Schulkind (2000) Experiment 2 that demonstrated the importance of instructing overlapping (response-relevant) stimulus categories on both tasks.

Experiments 4-5 (Chapter 5) extend the results obtained with the dual task approach by employing the same response-instruction logic to a Simon-like task similar to that used by Hommel (1995; 1996c), in which left and right keypress responses were arbitrarily mapped to centrally presented stimuli (letter identity). Irrelevant spatial information was provided by go/no-go signals (vertical or horizontal bars) at different locations, the orientation of which indicated whether the prepared response was to be executed or not. Hence, the Experiments presented in Chapter 5 extend the Simon et al. (1976) study by directly comparing the Simon effect (i.e., the impact of irrelevant spatial stimulus information) under spatial vs. non-spatial response instructions. Whereas both the ‘spatial only’ and the weak version of the direct coding hypothesis again predict that the spatial (Simon) effect should be largely unaffected by response instructions, the strong direct coding (i.e., intentional weighing) hypothesis predicts that the Simon effect should be reduced under non-spatial as compared to spatial response instructions.

In Experiments 1-3 (Chapter 4), mapping instructions were stated only once at the beginning of the experiments. Therefore, practice analyses were carried out to assess whether participants recoded instructed responses during the course of the experiments. In contrast, in Experiments 4 and 5 (Chapter 5) a procedural modification was introduced that allowed to deconfound the effects of instructions and practice present in all experiments so far (at least those that I am aware of). This was done by instructing new S-R mappings (i.e., new letter-response pairings) on each trial.

4 Dual Task Experiments

As noted in Chapter 3.3, the first goal of the set of experiments presented in this chapter was to extend the findings of inter-task consistency effects (e.g., Hommel, 1998; Lien & Proctor, 2000) to more abstract response dimensions, that is, to response instructions that do not encourage spatial coding. If participants indeed arbitrarily code their responses on two concurrently performed tasks as instructed – as predicted by (both versions of) the direct coding hypothesis – then R-R forward and backward compatibility of the kind observed by Hommel (1998) should generalize to more abstract response dimensions. Thus, instructions in the present Experiments 2 and 3 suggested color coding of manual responses instead of spatial coding. Manual responses cannot reasonably be assumed to be pre-defined with respect to color prior to instruction because no training phase with consistent color-effects was administered.

The second, related goal of the present set of experiments was to explore whether R-R (backward) compatibility effects and their reduction for non-overlapping tasks (Logan & Schulkind, 2000) depend on the overlap of instructed response representations. As discussed in Chapter 3.2 and 3.3, the experiments reported by Logan and colleagues are inconclusive with respect to whether overlap in instructed <u>response</u> categories is required for inter-task consistency effects to occur.

To address these questions, I used the experimental dual-task procedure of Hommel (1998, Experiment 1) involving a verbal and a manual task with arbitrary S-R mappings that had to be performed in close succession. I manipulated the overlap in response representations by varying the instructions of the manual keypress responses. More specifically, I instructed the left and the right response keys on the manual task either as *left* and *right* keys (Exp. 1) or as *blue* and *green* keys (Exp. 2 and Exp. 3), and required either “left” vs. “right” verbalizations (Exp. 1 and Exp. 3) or “blue” vs. “green” verbalizations (Exp. 2) on the verbal task. If indeed instructed response labels affect response coding, then one would expect that participants code their manual responses in terms of location under left/right instructions (Exp. 1), but that, under color instruction, conceptual color codes become part of the response representations and are or can be used in the control of manual responding (Exp. 2 and 3). This should lead to (forward and backward) compatibility effects on both the verbal and the manual task whenever the verbal task requires the same conceptual codes as the manual task, that is, when both the verbal and the manual task are coded in terms of location (Exp. 1) or color (Exp. 2).

In contrast, if response coding is restricted to spatial coding, as suggested by the spatial coding hypothesis, cross-task compatibility should be restricted to the spatial dimension (Experiment 1).

In addition, if such effects depend on the amount of overlap of task relevant (instructed) response dimensions, then they should be eliminated or extremely reduced when instructed response dimensions differ across tasks, that is, when the verbal task requires location coding, but instructions suggest color coding of left and right keypresses on the manual task (Exp. 3). In contrast, both the weak coding hypothesis that assumes comparable effects for implicit (uninstructed) as for explicit (instructed) overlap and the spatial coding hypothesis predict spatial inter-task consistency effects in Experiment 3 similar to those in Experiment 1.

For each experiment, additional practice analyses including the practice block were carried out. Although similar analyses performed by Hommel (1998) suggest that practice does not affect the size of backward (and forward) compatibility under spatial response instructions, these analyses were included here to assess whether subjects (spatially) re-code non-spatially instructed responses after practice. If so, one might expect potential color effects in Experiment 2 to decrease with practice, and a location-based effect to build up after some amount of practice in Experiment 3.

4.1 Experiment 1

Experiment 1 was a conceptual replication of Hommel (1998, Experiment 1). Participants were asked to perform a verbal and a manual task in close succession. As in the Hommel experiment, stimuli were univalent and were arbitrarily mapped to left and right verbalizations viz. keypresses in a 1:1 fashion. Experiment 1 differed from Hommel (1998, Exp. 1) in two aspects. First, in the present experiment the verbal task was the primary task, and the manual task was the secondary task. The order of the tasks was reversed to allow assessment of backward effects from the manual task. As has been argued in Chapter 3.2, backward compatibility effects not only show that the overlapping codes are part of the Task 2 response representation, but indicate that these codes are actively used to access and guide responding on the secondary task. Second, I used non-integrated stimuli for the two tasks, namely tone stimuli and geometric form stimuli, which were presented asynchronously with a SOA of 50 ms. This was done to reduce the probability that participants recode the instructed S-R rules by assigning one response code to an integrated stimulus representation on compatible trials (e.g., red H <> right).

I expected to obtain R-R compatibility effects between the responses on both tasks. Hence, both verbal and manual responses were expected to be faster on response-compatible (e.g., a verbal “left” response followed by a left keypress) than on response-incompatible trials (e.g., a verbal “left” response followed by a right keypress).

4.1.1 Method

Participants

Forty-seven undergraduate students (28 female, 19 male, mean age = 24 years) at Humboldt University, Berlin, participated for partial fulfillment of course credit. All subjects had normal or corrected to normal vision. Thirteen subjects of the total sample were excluded from the data analyses according to pre-defined criteria (see Section Data Analysis for a description of the final sample).

Apparatus and Stimuli

The experiment was controlled by Pentium II computers with SoundBlaster 16 Audio cards that were connected to external speakers, headphones with microphones attached to the headphones and a standard (German) keyboard. Stimulus presentation and response recording was controlled by a modified version of the Runword software, a freeware provided by C.T. Kello. Runword runs in DOS mode on IBM compatible PCs with (ISA compatible) SoundBlaster 16 Audio Cards.

Two different tone stimuli were used in the verbal task. Tone stimuli were generated by converting two different Windows wav-files (krt08.wav and schl05.wav, cut down to 50 ms duration each) into voc-files. Krt08 is essentially a squeak tone, whereas schl05 can best be described as a snap tone. Tone stimuli were displayed by Runword via speaker output. The speakers were located to the left and the right of the screen, and tones were simultaneously presented through both speakers. Volume was adjusted individually before the experiment started.

Squares and circles served as visual stimuli on the manual task. The diameter of the circle was 3 cm, as was the length of the sides of the square. Visual stimuli appeared as black frames against a white background plate of 10 x 8 cm (height x width) size at the screen center of 17 inch monitors. The viewing distance was approximately 50 cm.

In order to record vocal responses at a standardized level, the microphone was calibrated before the experiment started. During the experiment, Runword generated voc-files for each verbal response. Verbal RTs were determined after the experiment by a software voice-key provided by Runword, applying a fixed threshold as response criterion. Response alternatives on the verbal task were "links" and "rechts," the German words for "left" and "right." The left and the right keys on the manual task were the 'y' and the '.' keys on a standard German keyboard, respectively.

Design and Procedure

Each session started with a written and verbal instruction specifying the required S-R mappings and describing the sequence of events on each trial. Tone stimuli were labeled "Knackton" and "Quietschton," the German words for snap tone and squeak tone, and were demonstrated during instruction. Verbal responses were assigned to tone stimuli, and manual responses to form stimuli. The four different mapping combinations (of tone and form stimuli to verbal and manual responses, respectively) were approximately counterbalanced across participants. Compatibility was defined in terms of overlap in 'response location,' so for example, a verbal "links" response followed by a left keypress was considered compatible. After instruction, participants were required to recall the mappings correctly and then received a written reminder of the instructed mapping rules that was accessible during the entire course of the experiment.

Participants then worked through eight blocks of 32 trials each. Each block contained 8 replications of each of the four combinations of tone and form stimuli, amounting to 16 trials of each compatibility condition per block. The first block was treated as a practice block and was not considered in the main analyses, resulting in 7 experimental blocks with a total of 56 trials per stimulus combination (112 trials per compatibility condition) overall. However, the practice block was considered in the practice analyses. More specifically, for the practice analyses, two blocks each (including the practice block) were aggregated into a block-cluster, resulting in four block-clusters that provided the basis to test for practice effects. Accordingly, each block-cluster contained 16 replications of each of the four tone-form stimulus combinations, amounting to 32 trials in each compatibility condition per block-cluster and 128 compatible and incompatible trials overall.

Trials were presented in one of four different quasi-random orders that required stimulus combinations to follow each other about equally often. These quasi-random trial sequences were determined before the experiment and were counterbalanced across participants.

Each trial started with a plate saying “Fertig?” (‘Ready?’) presented for at least 1000 ms. When participants were ready to commence the trial, they pressed the space key to initiate the trial. After a 700 ms blank interval, a fixation cross appeared. Seven hundred ms later one of the tone stimuli was displayed for 50 ms while the fixation cross remained on the screen. Simultaneously with the offset of the tone, one of the form stimuli replaced the fixation cross and remained on the screen for 1200 ms. Response recording for both the verbal and the manual task was initiated at the onset of the form stimulus and terminated after 2500 ms had passed. Participants had to respond verbally to tone stimuli first, and then to press the left or right key with their left and right index fingers in response to the form stimuli. Instructions emphasized the requirement to perform the two responses in strict serial order, and participants were reminded of the required response order after the practice block.

4.1.2 Results

Data Analysis

Main Results and Additional Analyses. Data from 13 participants were excluded from the main analyses of Experiment 1 because they did not perform at a pre-experimentally determined level of performance on one or both of the tasks[7]. Nine participants did not comply with instructions in that they performed the manual before the verbal task on more than 10% of the trials on the seven experimental blocks. Three additional participants were excluded because they produced more than 30% errors, misses, and/or uncodable vocal responses, again across the seven experimental blocks. The latter exclusion criterion was applied in order to guarantee reliable RT results[8]. One additional participant had to be excluded due to recording problems of vocal responses. As a consequence, data from thirty-four undergraduate students (20 female, 14 male, mean age = 24 years) entered into the main analyses of Experiment 1.

---

[7] Actually, the total sample (N = 47, see Section Participants, above) contains participants that were excluded according to these criteria and subjects that were tested in their place. Therefore, mappings and quasi-random sequences were approximately counterbalanced in the remaining (N = 34) sample. This also holds for Experiments 2 and 3.

[8] Separate analyses on the data of excluded participants revealed a similar pattern of results as in the original analyses.

In the reduced sample, response order errors occurred on 1.0 % of the experimental trials and were excluded from analyses. Trials on which one or both of the responses was faster than 50 ms or slower than 2000 ms accounted for 1.1% of the trials and were also excluded from further analyses, as were double error trials in which responses on both tasks were incorrect or omitted (1.0% of the trials) because double errors are difficult to interpret (cf. Schuch & Koch, submitted).

For the remaining data, median RTs for trials on which both responses were correct, and the percentage of invalid trials (PI) including errors and response omissions on only one task were computed for each task as a function of the factors under consideration in the main analyses and the additional analyses, respectively. The decision to go with median RTs instead of mean RTs was motivated by the consideration that medians are less sensitive to outliers than means unless individual raw data are trimmed (cf. Ratcliff, 1993), and data trimming for individual subjects would have led to different criteria in the main analyses and the practice analyses that included the practice block in the analysis (see below). However, analyses on (untrimmed) means that were run for reasons of comparability indicated a very similar pattern of results in all three experiments, and hence lead to the same conclusions.

Because response recording of verbal responses was initiated 50 ms after the onset of the tone stimuli (see method section, above), a constant amount of 50 ms was added to all verbal RTs before data screening, data aggregation, and analysis. This was also done in Experiments 2 and 3, as well as in the practice analyses. For all factors including more than two within-subject conditions, reported p-values are based on Greenhouse-Geisser corrected degrees of freedoms. This also applies to the practice analyses and the other experiments presented in this thesis.

Practice Analyses. For the practice analyses, the data, including those from the practice block (which are not considered in the main analyses), were aggregated into four block-clusters, each block-cluster containing data from two blocks. Block-clusters instead of experimental blocks were considered in order to smooth the learning curves and to avoid unreliable Block 1 estimates resulting from high error rates in the first (practice) block.

According to the pre-experimentally defined exclusion criteria described above, two additional participants had to be excluded from the practice analysis because they produced more than 30% invalid trials overall when the practice block was included, leaving $N = 32$ participants (19 female, 13 male, mean age = 23.6 years) in the practice analyses.

In the final sample, response order errors accounted for 1.1% of the trials and were excluded from the analysis. Trials with responses faster than 50 ms or slower than 2000 ms occurred on 1.3% of the trials and were excluded, too, as were double error trials in which responses on both tasks were incorrect or omitted (1.6% of the trials).

For the remaining data, median RTs for trials on which both responses were correct, and PIs (again only considering trials with errors or omissions on one task) were computed for each task as a function of block-cluster (1-4), task (verbal, manual), and compatibility (compatible, incompatible).

Main Results

For the standard analyses, median RTs for trials on which both responses were correct, and PIs including errors and response omissions on only one task were computed for each task as a function of compatibility between verbal and manual responses. Table 5 shows the group means of the individual median RTs and of the accuracy data across experimental blocks for the reduced sample ($N = 34$).

Table 5. Mean Median Reaction Times (RT, in ms) and % Invalid (PI) for Verbal (Primary) and Manual (Secondary) Responses as a Function of Response-Response Compatibility in Experiment 1.

	Compatible		Incompatible		Δ[a]	
Response	RT	PI	RT	PI	RT	PI
Verbal	553 (565)[b]	2.0 (2.1)	601 (590)	4.4 (4.4)	48 (25)	2.4 (2.3)
Manual	936 (931)[b]	2.5 (2.5)	982 (969)	4.5 (4.4)	46 (38)	2.0 (1.9)

Note: [a] Columns labeled Δ indicate effect sizes of the compatibility effects (incompatible minus compatible).
[b] Numbers in parentheses represent results based on $n = 33$ participants, excluding the compatibility effect outlier (see text for details).

RT. Median RTs were submitted to an analysis of variance (ANOVA) with task (verbal, manual) and response compatibility (compatible, incompatible) as within-subjects factors. This analysis yielded a significant main effect of task, $F(1,33) = 329.26$, $p < .001$, $MSe = 15{,}048.23$, indicating faster responses on the verbal task. More important for the present purposes, the main effect of compatibility also reached significance, $F(1,33) = 6.61$, $p < .05$, $MSe = 11{,}451.85$, whereas the interaction between task and compatibility did not, $F(1,33) < 1$, $MSe = 2{,}936.96$, suggesting that the verbal (primary) and the manual (secondary) tasks were similarly affected by the compatibility relation between responses. However, planned comparisons that tested the effect of compatibility separately for each task, showed that compatibility only reached significance in the analysis of manual responses,

$F(1,33) = 9.19$, $p < .01$, $MSe = 8{,}011.04$. It just missed significance for the verbal task, $F(1,33) = 3.75$, $p < .07$, $MSe = 20{,}766.58$. A closer inspection of the verbal data revealed that this outcome was due to a single participant who showed a particularly large compatibility effect on the verbal task (> 800 ms). Eliminating this participant from the analyses numerically reduced the compatibility effects for verbal and manual responses to 25 ms and 37 ms (see Table 5, numbers in parentheses), respectively, but now the verbal effect also reached significance, $F(1,32) = 6.8$, $p < .05$, $MSe = 1{,}503.68$, while all other effects (including those in the omnibus analysis) remained qualitatively the same.

PI. The ANOVA on PIs[9] yielded only a significant main effect of compatibility, $F(1,33) = 13.95$, $p < .001$, $MSe = 11.72$; the main effect of task and the interaction between task and compatibility were not significant (both F's < 1). Planned comparisons that tested the compatibility effect separately for each task revealed a significant compatibility effect for manual responses, $F(1,33) = 6.4$, $p < .05$, $MSe = 21.16$, and a significant effect for verbal responses, $F(1,33) = 15.3$, $p < .001$, $MSe = 12.69$.

Additional Analyses

In order to assess whether the backward compatibility effect on the verbal task depended on response grouping on a certain number of trials, that is, on withholding R1 until R2 is selected, I additionally assessed whether the compatibility effects depended on the lag between R1 and R2. If the compatibility effects were due to response grouping, then they should be particularly pronounced at short inter-response intervals (IRIs) and decrease with increasing IRI. Moreover, verbal RTs should be slowed at short as compared to long IRIs. For each participant and compatibility condition, I therefore determined IRI quintiles and calculated median RTs for verbal and manual responses for each quintile. Figure 3 shows the group means of the medians for each IRI quintile.

Of particular interest for an interpretation of the backward effect is how verbal responses behave across IRIs. As is clear from inspection of Figure 3, verbal RTs remain relatively constant across IRIs. More importantly, the size of the compatibility effect in verbal responses seems unaffected by IRI.

[9] Excluding the participant that produces compatibility effect outlier in RTs from the PI analysis did not affect the error rates in any way (see Table 5). Therefore I conducted the PI analysis on the complete data set.

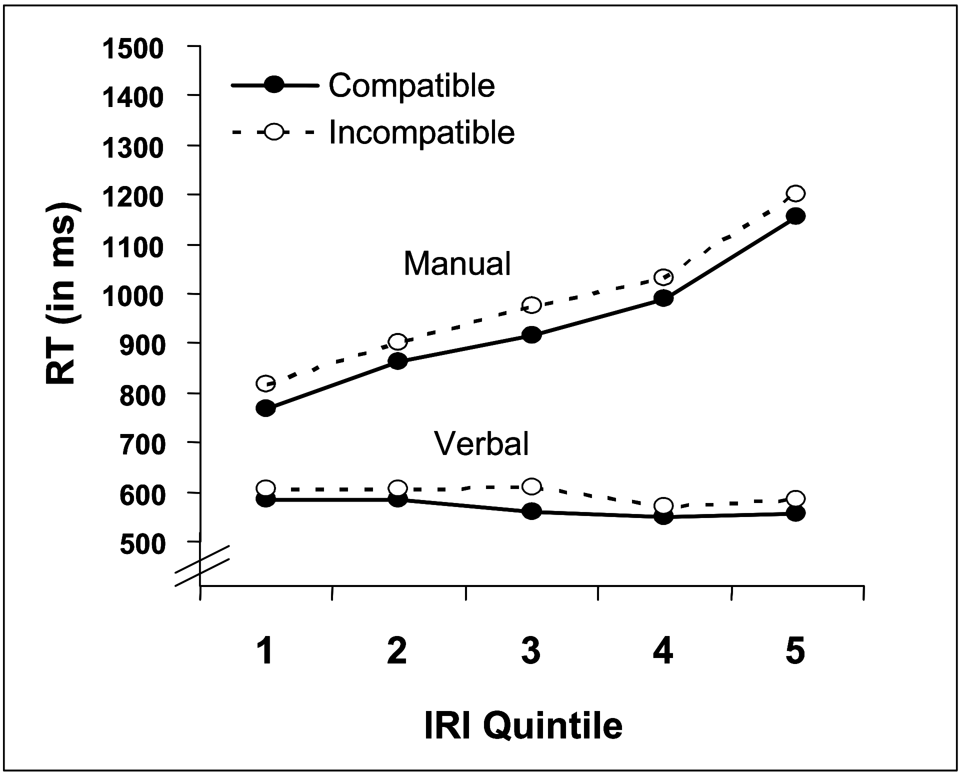

Figure 3. Mean median verbal and manual reaction times (RTs) as a function of inter-response-interval (IRI) and compatibility between verbal and manual responses in Experiment 1. Mean median IRI quintiles were 191, 279, 354, 435, and 573 ms for compatible trials, and 220, 295, 359, 456, and 598 ms for incompatible trials.

These conclusions were supported by a two-way ANOVA that produced a significant main effect of compatibility, $F(1,33) = 7.45$, $p < .05$, $MSe = 9{,}238.59$; the main effect of IRI quintile and the interaction between IRI and compatibility did not reach significance (both F's < 1). The analogous analysis on manual responses that was run for reasons of comparability, showed that manual RTs increased with IRI quintile, $F(4,132) = 156.83$, $p < .001$, $MSe = 9{,}100.94$, and also revealed a significant main effect of compatibility, $F(1,33) = 11.25$, $p < .01$, $MSe = 16{,}226.32$. However, the interaction of quintile and compatibility did not reach significance, $F(4,132) < 1$, $MSe = 3{,}645.05$, implying that the size of the manual compatibility effect was independent of IRI.

In addition, in order to gain insight into the temporal dynamics of the compatibility effects, distribution analyses were carried out on the RT data. To this end, RT quintiles were determined for each participant, task, and compatibility condition, and median RTs were computed accordingly. Figure 4 shows the resulting group means of the individual medians for each RT quintile and condition. Note that averaging the means of medians across quintiles leads to higher mean median scores per condition than those presented in Table 5. Whereas such a difference would not be expected in analyses on means, it is not particularly surprising with respect to medians. This is so because medians are more sensitive to the positive skewness of RT distributions than are means (i.e., they respond less strongly to data in the slow end tail of the RT distribution), leading to higher overall RT estimates in the quintile analysis that "weighs" responses in the slow end tail of the distribution more heavily. Therefore, the tabled values (i.e., mean median overall RTs) roughly correspond to the mean medians of the third RT quintile of each condition, instead of the means across quintiles.

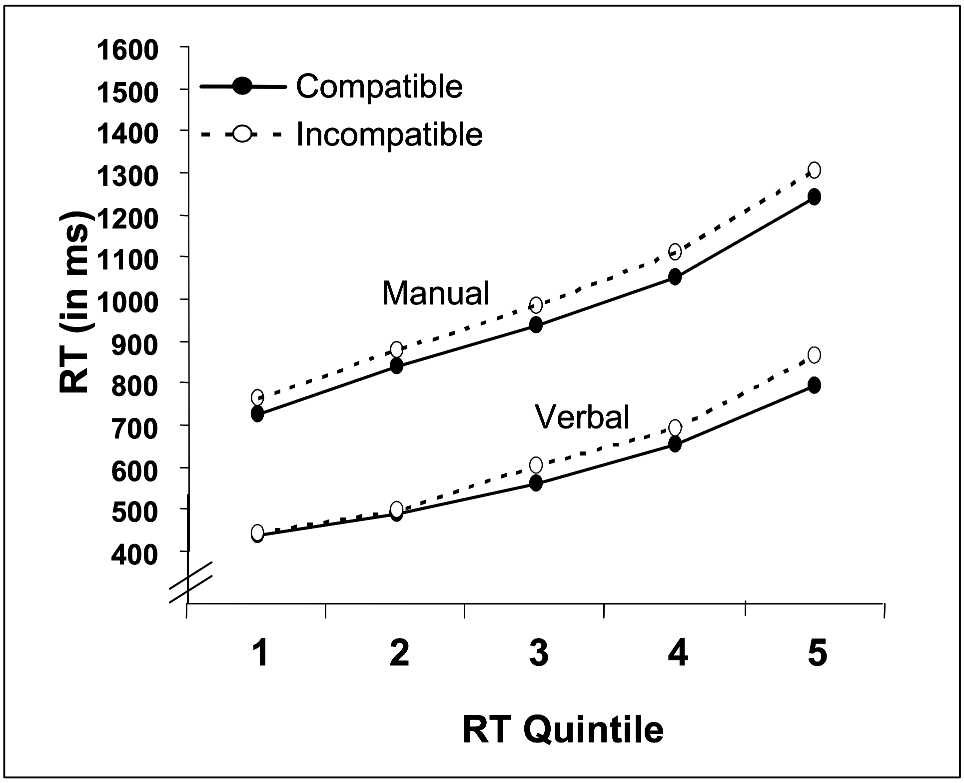

Figure 4. Mean median verbal and manual reaction time quintiles (RT quintiles) as a function of compatibility between verbal (primary) and manual (secondary) responses in Experiment 1.

The calculated median RTs were entered into a 2 (task) x 2 (compatibility) x 5 (RT quintile) ANOVA. Aside from producing a trivial main effect ($F(1,132) = 195.57$, $p < .01$, $MSe = 22{,}563.13$), quintile participated in both two-way interactions, $F(4,132) = 13.92$, $p < .01$, $MSe = 5{,}600.32$ for the interaction of quintile and task, and $F(4,132) = 6.03$, $p < .01$, $MSe = 2{,}181.77$, for the interaction of RT quintile and compatibility. The three-way-interaction of task, quintile, and compatibility did not reach significance, $F(4,132) = 1.35$, $p > .26$, $MSe = 1{,}244.16$, indicating that the compatibility effects on the two tasks similarly increased with increasing RT.

Practice

Although additional analyses performed by Hommel (1998) suggest that practice does not affect the size of spatial backward (and forward) compatibility, additional analyses were performed on the Experiment 1 data to assess whether this was also true for the present experiment. To this end, median RTs for trials on which both responses were correct, and the percentage of invalid trials on only one task were computed for each task as a function of block-cluster (1-4), task (verbal, manual), and compatibility (compatible, incompatible) on the basis of the data of the remaining thirty-two participants (see Section Data Analysis)[10], the group means for which are presented in Table 6. Note that the means in the columns of Table 6 labeled "overall" do not correspond to the values presented in Table 5 because (a) the data from the practice block are included, (b) the data are based on a reduced sample, and (c) because aggregating medians in a two-step fashion (i.e., first on block-clusters and then across block-clusters) can lead to different overall means even if the same data are included.

Table 6. Mean Median Reaction Times (RT, in ms) and % Invalid (PI) for Verbal (Primary) and Manual (Secondary) Responses as a Function of Response-Response Compatibility and Block Cluster in Experiment 1.

		Block Cluster									
		1		2		3		4		overall	
Response	R1-R2	RT	PI	RT	PI	RT	PI	RT	PI	RT	PI
verbal	compatible	638	4.6	592	2.1	556	1.2	540	1.7	582	2.4
	incompatible	657	8.7	626	4.4	588	3.4	556	3.9	607	5.1
	Δ	19	4.1	34	2.3	32	2.2	16	2.2	25	2.7
manual	compatible	1069	6.5	973	2.3	919	2.1	893	2.1	963	3.3
	incompatible	1118	8.3	1007	5.5	960	2.9	917	3.0	1000	4.9
	Δ	49	1.8	34	3.2	41	0.8	24	0.9	37	1.6

Note: Rows labeled Δ indicate effect sizes of the compatibility effects. The two rightmost columns contain the means across blocks.

RT. Median RTs were submitted to an ANOVA with task (verbal, manual), response compatibility (compatible, incompatible), and block-cluster (1-4) as within-subjects factors. This analysis again yielded a significant main effect of task, $F(1,31) = 359.91$, $p < .001$, $MSe = 53{,}464.38$, indicating faster responses on the verbal task. The main effect of block-cluster was also significant, $F(3,93) = 34.44$, $p < .001$, $MSe = 14{,}552.6$, as was the interaction

[10] The outlier subject from the main analysis did not fulfill the pre-defined error criteria when the practice block was included and was therefore excluded from the practice analyses.

between task and block-cluster, $F(3,93) = 11.19$, $p < .001$, $MSe = 4{,}584.19$, implying that, although both tasks benefited from practice, the reduction of RT with practice was more pronounced for the manual task. As in the analysis without block-cluster as a factor (see Section Main Results, above), the main effect of compatibility was significant, $F(1,31) = 15.33$, $p < .001$, $MSe = 8{,}051.67$, whereas the interaction of task and compatibility was not, $F(1,31) = 1.29$, $p > .26$, $MSe = 3{,}483.72$.

More importantly, block-cluster did not interact with compatibility. That is, neither the interaction of block-cluster and compatibility ($F(3,93) < 1$, $MSe = 5{,}048.3$), nor the three-way interaction of block-cluster, compatibility and task ($F(3,93) < 1$, $MSe = 1{,}326.99$) reached significance, indicating that practice, while speeding up overall RTs, did not affect the compatibility effects. The latter conclusion receives support from separate analyses for each task that did not yield significant block-cluster x compatibility interactions either (both *F*'s < 1).

PI. The ANOVA on PIs only yielded significant main effects of compatibility, $F(1,31) = 10.26.95$, $p < .01$, $MSe = 59.57$, and of block-cluster, $F(3,93) = 23.31$, $p < .001$, $MSe = 24.95$. Block-cluster did not interact with any other variable (all *F*'s ≤ 1). As in the RT analysis, the interaction of task and compatibility did not reach significance ($F(1,31) = 1.6$, $p > .15$, $MSe = 22.08$), but unlike RTs there was no overall difference between tasks regarding PIs ($F(1,31) < 1$, $MSe = 72.08$). Aside from this relatively minor discrepancy, the error results closely resemble the RT results.

4.1.3 Discussion

In Experiment 1, I found that both primary and secondary task performance was influenced by the compatibility relation between the responses on a verbal and a concurrently performed manual task. More specifically, both verbal (primary) and manual (secondary) responses were faster and less error prone when the same response code was required on the two tasks (e.g., when both responses were left) than when correct responding required different spatial codes (i.e., when one response was left and the other was right). Hence, I was able to replicate the Hommel (1998) findings, although, in my Experiment 1, the size of the compatibility effects on the two tasks was numerically smaller than in the Hommel experiment. Possibly, the reduced size of the compatibility effects was due to the use of non-integrated stimuli and to asynchronous stimulus presentation. The former may have reduced the likelihood of stimulus- or response recoding into integrated rules on compatible trials, whereas the latter may have reduced the overlap in S-R processing of the two tasks.

A second difference between the Hommel experiment and my Experiment 1 (but see Experiment 2 below) is that in the former, but not in the latter experiment, the Task 2 (forward) effect was significantly larger than the backward effect. Whereas Hommel (1998) explained his finding by some additional contribution from meaning based forward-priming due to less-than-perfect response-reset, Hommel and Eglau (2002) noted that in many studies (using a slightly different paradigm) "compatibility effects on the secondary task commonly reflect little more than mere propagation of the effect from the primary task" (p. 272). At present, I do not have a good explanation for why some studies find comparable effects on both tasks while others do not (but see Hommel & Eglau, 2002, for an attempt). However, because my primary concern is with backward compatibility, this question does not appear to be of primary importance.

Apart from these differences, the present results are similar to those obtained by Hommel (1998), in that the backward compatibility effect was more pronounced for long verbal RTs, indicating that compatibility effects are larger when the overlap between S1-R1 and S2-R2 processing is enhanced. However, the effect in verbal responses did not depend on IRI, implying that response grouping cannot explain the backward compatibility effect.

Moreover, the effects did not depend on practice, that is, they remained relatively constant across blocks although overall RTs decreased with practice. This finding presumably implies that – whatever changes regarding processes involved in translation (e.g., shortcuts or instance based processing) lead to more efficient translation with practice – the codes responsible for the compatibility effects remain part of the response representations.

Taken together, these results have three important implications. First, when responses on two concurrently performed tasks are instructed in terms of location, then the same abstract conceptual (left and right) codes mediate responding on the two tasks. Second, S-R translation for the two tasks must have overlapped in time to some degree. Third, response coding remains largely unaffected by practice, at least within the range of practice used in the present experiment.

However, Experiment 1 cannot discriminate between the different hypotheses regarding how instructions influence response coding. As discussed earlier (see Chapter 3.2), this is so because the instructions in my Experiment 1, just like the instructions by Hommel (1998), Lien and Proctor (2000), and by Koch and Prinz (submitted) encouraged spatial coding on the two tasks. Therefore the results are consistent with both, the direct and the obligatory spatial coding hypothesis.

In Experiment 2, an attempt was made to discriminate between the spatial coding hypothesis, on the one hand, and (both versions of) the direct coding hypothesis, on the other hand. This was done by instructing non-spatial response dimensions on the two tasks.

4.2 Experiment 2

In Experiment 2, left and right response keys on the manual task were no longer instructed as *left* vs. *right*, but as *blue* vs. *green*. The verbal task also required “blue” vs. “green” responses.

If color instructions lead to integration and use of color codes in manual response representations and responding, as predicted by both versions of the direct coding hypothesis, then color based R-R compatibility effects between verbal and manual color responses should be observed (e.g., a “blue” verbalization followed by a blue keypress should be easier than, say, a “blue” verbalization followed by a green keypress). Alternatively, if manual responses are spatially coded in an instruction-independent way, no such color-effect should be obtained.

Note that instructing arbitrary S-R mappings in terms of (response) color differs from the ‘standard’ Hedge and Marsh task (see Chapter 3.1.4) in that the relevant stimulus attribute in the latter is color, but an arbitrary stimulus attribute was mapped to color-responses in my Experiment 2. Whereas participants can perform the H&M task on the basis of a ‘same’ or ‘opposite’ rule, the use of such a simple rule is not possible for arbitrary S-R mappings.

As has been argued in Chapter 3, the strongest evidence for arbitrary coding of <u>manual</u> keypress responses so far comes from the response-effect literature (cf. Chapter 3.1.3). For instance, Hommel (submitted) demonstrated that consistently presented color effects become integrated into the action representations of left and right keypress responses, and can be used to access and guide responding, as indicated by color-based S-R compatibility effects observed when color frames served as distractors in the test phase.

In a similar vein, Koch and Kunde (in press) demonstrated that <u>verbal</u> color responses (i.e., verbalizing color names in response to digits) were affected by the identity of color words presented after responding. Response-effect words were either congruently colored or not colored (e.g., the word BLUE was presented in blue or grey color). Responses were faster when the color words were consistent with the response (e.g., a “green” response followed by the word GREEN) than when incompatible color words (e.g., a “green” response followed by

BLUE) served as response effects, and response-effect compatibility was even more pronounced when the color words were congruently colored. This finding suggests that conceptual color codes mediated verbal response selection.

The crucial differences between these experiments and the present Experiment 2 are that in the present experiment (a) there was no practice phase in which color effects might have become associated with manual responses, as was the case in the Hommel (submitted) experiment, and (b) no response-compatible or incompatible action effects were presented upon responding. Rather, I simply instructed participants to press the blue or green key in response to form stimuli. Whereas stimulus-to-color mapping remained constant throughout the experiment, color-to-(left/right) key assignment varied unpredictably from trial to trial.

4.2.1 Method

Participants

Forty-two undergraduate students (31 female, 11 male, mean age = 23.2 years) at Humboldt University, Berlin, received either € 7,- or partial course credit for participation[11].

Apparatus and Stimuli

Apparatus and stimuli, as well as response keys and response recording were identical to those used in Experiment 1.

Design and Procedure

In Experiment 2, I instructed left vs. right keypresses as blue vs. green keypresses and also required blue vs. green verbalizations on the verbal task. Accordingly, participants were asked to first respond verbally to tone stimuli by saying either “blau” or “grün” (the German words for blue and green, respectively), and then to press either the blue or the green key, depending on the form stimulus. The four different stimulus-to-color mapping combinations (e.g., combination 1: squeak tone < > “green”, snap tone < > “blue”; square < > green key, circle < > blue key; combination 2: squeak tone < > “blue” …; square < > green key …) were counterbalanced across participants and remained constant throughout the experiment for in-

[11] As in Experiment 1, these data describe the total sample, including the participants that were excluded and those that were tested in their place (see data analysis section for details).

dividual participants. Congruency was defined in terms of 'response color' (for example, a "blue" verbalization followed by a blue keypress was considered congruent).

Because the keys were not colored themselves and instruction only specified the stimulus-to-color mapping for the manual task, leaving open the color-to-(left/right) key assignment on a given trial, each trial started with the presentation of a plate that graphically depicted the color-to-key assignment for the next trial. For instance, when the plate showed a green color patch on the left and a blue color patch on the right (with pictograms of a left and a right hand beneath the color patches), then participants knew that the green key would be the key on the left and the blue key the one on the right on the forthcoming trial. The color plate remained on the screen for at least 1000 ms after which participants could trigger stimulus presentation by pressing the space key.

Color-to-key assignment varied randomly from trial to trial with the following constraints. First, in each block, each combination of tone and form stimuli appeared equally often under each color-to-key assignment. Second, each color-to-key assignment x congruency condition followed each other about equally often. Four quasi-random sequences (i.e., trial orders) fulfilling these constraints were again determined before the experiment, and were counterbalanced across participants.

The first block again served as a practice block and was not considered in the main and additional analyses. On the remaining seven experimental blocks, participants saw a total of 112 trials in each congruency condition, 28 in each stimulus combination x color-to-key assignment condition. Again, the practice block was included in the practice analyses. Hence, each block-cluster (on which the practice analyses are based) contained 16 congruent and incongruent trials under each color-to-key assignment, realized by 8 replications of each S1-S2 combination.

The rest of the procedure was identical to the procedure of Experiment 1.

4.2.2 Results

Data Analysis

Main Results and Additional Analyses. Two participants were excluded from Experiment 2 because they performed the manual task before the verbal task on more than 10% of the trials, and further six participants were excluded because they produced more than 30% errors, misses and uncodable vocal responses overall for the same reasons as in Experiment 1.

Additional participants were tested in their place. As in Experiment 1, the pattern of results for excluded participants was similar to that for included participants.

Hence, only data of a reduced sample consisting of thirty-four undergraduate students (26 female, 8 male, mean age = 22.7 years) were analyzed. As in Experiment 1, median RTs for trials on which both responses were correct, and PIs for trials with only one invalid response were computed for each participant as a function of the variables under consideration (see Sections Main Results and Additional Analyses, below). This was again done after excluding trials with response order errors (0.93%), those with responses faster than 50 ms or slower than 2000 ms (2.32%), as well as trials with double errors (0.83%).

Practice Analyses. Based on the exclusion criteria, two additional participants had to be excluded because they produced more than 30% invalid trials overall when the data from the practice block were included, leaving $N = 32$ participants (24 female, 8 male, mean age = 22.6 years) in the practice analyses.

In the remaining data, response order errors occurred on 0.9% of the trials. These trials were excluded from the analyses, as were double errors (1.7% of the trials) and trials with responses that were faster than 50 ms or slower than 2000 ms on one or both tasks (2.4%). As in Experiment 1, median RTs for completely correct trials and PIs for trials with invalid responses on only one task were computed as a function of task, block-cluster (consisting of two blocks each, including the practice block), and R-R compatibility, to assess whether practice modifies the compatibility effects.

Main Results

Based on the data of $N = 34$ participants, median RTs and PIs were computed for each participant, task, and compatibility condition (see Table 7 for group means). The pattern of results can be described as follows. Manual responses seem to be slower and more error prone than verbal responses. More important for the present purposes, R-R compatibility affects the speed of responses as well as the likelihood to make an error or to omit a response on both tasks, although the compatibility effect seems to be somewhat larger on the (secondary) manual task.

Table 7. Mean Median Reaction Times (RT, in ms) and % Invalid (PI) for Verbal (Primary) and Manual (Secondary) Responses as a Function of Response-Response Compatibility in Experiment 2.

	Compatible		Incompatible		Δ	
Response	RT	PI	RT	PI	RT	PI
Verbal	656	2.0	692	4.0	36	2.0
Manual	1039	4.0	1114	6.5	75	2.5

Note: Columns labeled Δ indicate effect sizes of the compatibility effects (incompatible minus compatible).

These observations were supported by the ANOVA results.

RT. The omnibus ANOVA including task and compatibility as within-subjects factors revealed highly significant main effects of task, $F(1,33) = 291.77$, $p < .001$, $MSe = 18, 840.94$, and compatibility, $F(1,33) = 24.52$, $p < .001$, $MSe = 4,234.04$, as well as a highly significant interaction between task and compatibility, $F(1,33) = 21.69$, $p < .001$, $MSe = 585.72$. However, planned comparisons that tested the compatibility effects separately for the two tasks showed that compatibility was significant for both manual responses, $F(1,33) = 26.92$, $p < .001$, $MSe = 7,036.67$, and verbal responses, $F(1,33) = 16.8$, $p < .001$, $MSe = 2,620.84$.

PI. The PI analysis yielded significant main effects of task, $F(1,33) = 10.91$, $p < .01$, $MSe = 15.16$, and compatibility, $F(1,33) = 18.83$, $p < .001$, $MSe = 8.99$; the interaction between task and compatibility did not reach significance, $F(1,33) < 1$, $MSe = 6.56$. Planned comparisons of the compatibility conditions on the two tasks again revealed significant effects for both manual, $F(1,33) = 11.54$, $p < .01$, $MSe = 18.72$, and verbal responses, $F(1,33) = 10.36$, $p < .01$, $MSe = 12.39$.

Additional Analyses

To assess whether the backward compatibility effect depended on IRI, I again determined the IRI quintiles for each participant, task, and compatibility condition. Figure 5 shows the mean median RTs per task, compatibility condition, and IRI.

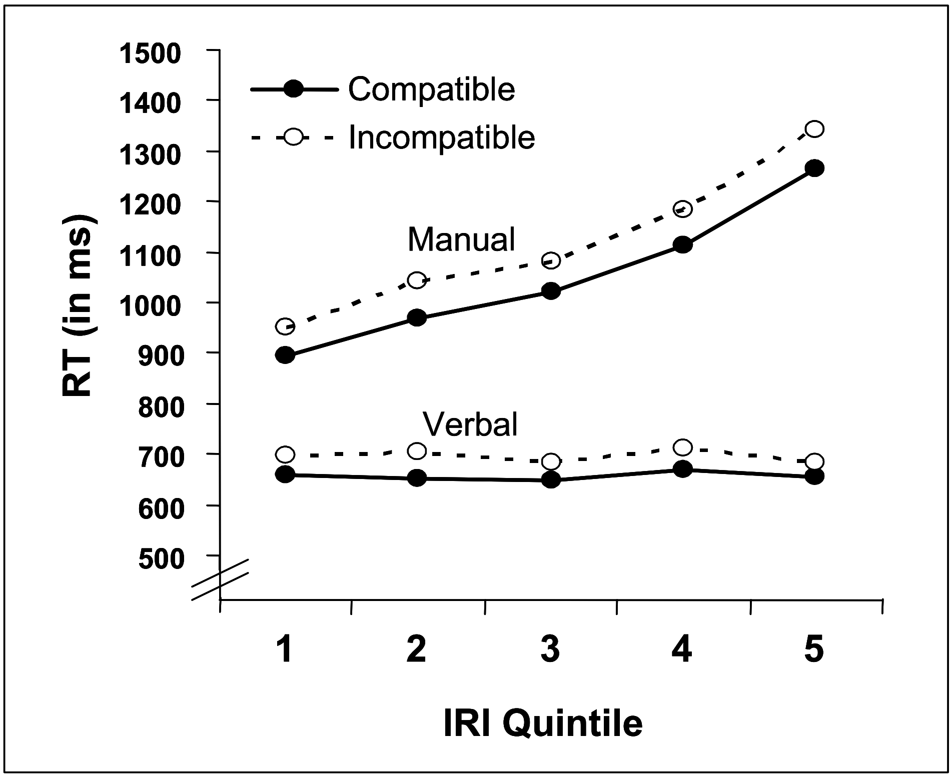

Figure 5. Mean median verbal and manual RTs as a function of IRI and compatibility between verbal and manual responses in Experiment 2. Mean median IRI quintiles were 234, 314, 372, 437, and 581 ms for compatible trials, and 261, 338, 398, 475, and 610 ms for incompatible trials.

Again, verbal RTs remained relatively constant across IRIs, and the size of the compatibility effect for verbal responses was unaffected by IRI, as confirmed by a two-way ANOVA that yielded a highly significant effect of compatibility, $F(1,33) = 17.7$, $p < .001$, $MSe = 7{,}282.61$, but no effect of IRI quintile, $F(4,132) < 1$, $MSe = 7{,}811.09$, and no significant interaction between compatibility and IRI quintile, $F(4,132) < 1$, $MSe = 2{,}722.36$. As in Experiment 1, manual RTs increased with increasing IRI, $F(4,132) = 154.39$, $p < .001$, $MSe = 9{,}327.74$, and also revealed a relatively constant compatibility effect across IRI, as reflected by a significant main effect of compatibility, $F(1,33) = 32.71$, $p < .001$, $MSe = 12{,}135.17$, and a nonsignificant interaction between IRI and compatibility, $F(4,132) < 1$, $MSe = 3{,}208.14$.

Distribution analyses on RT quintiles (see Figure 6), including task, compatibility, and RT quintile as within subjects factors, again revealed that quintile was significant, $F(4,132) = 429.72$, $p < .001$, $MSe = 11{,}271.06$, and that it interacted with task, $F(4,132) = 38.88$, $p < .001$, $MSe = 2{,}662.94$, as well as with compatibility, $F(4,132) = 15.92$, $p < .001$, $MSe = 1{,}463.62$. The three-way-interaction between RT quintile, task, and compatibility did not reach significance, $F(4,132) < 1$, $MSe = 504.13$, again indicating that the compatibility effects on both tasks increased as response speed decreased.

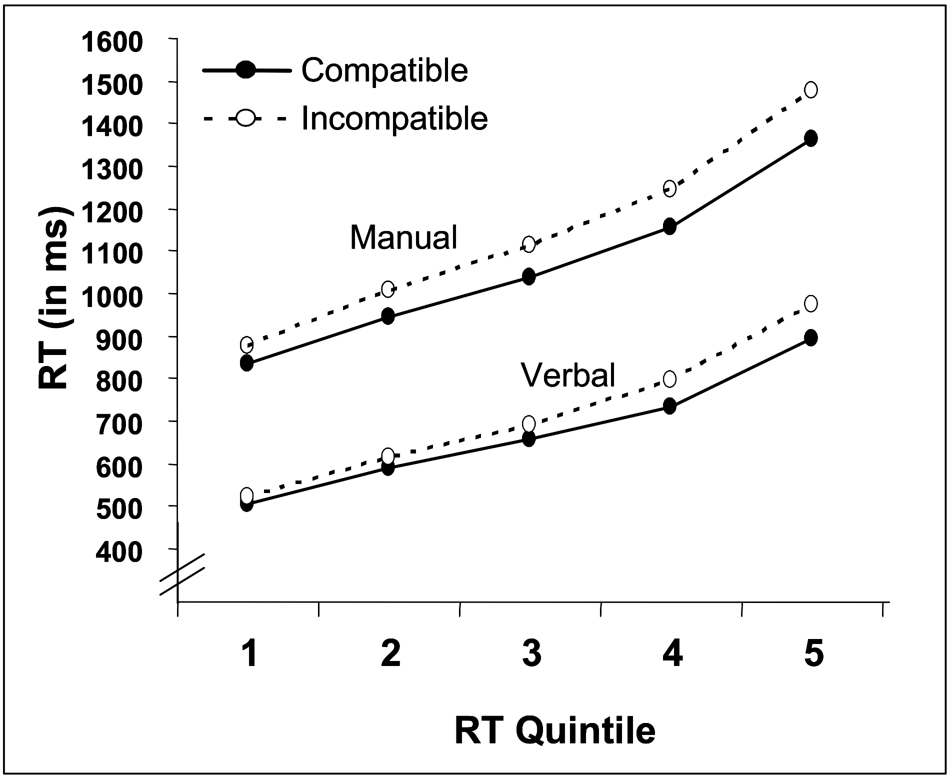

Figure 6. Mean median verbal and manual RT quintiles as a function of compatibility between verbal (primary) and manual (secondary) responses in Experiment 2.

Practice

As in Experiment 1, in order to assess the impact of practice on the compatibility effects, median RTs for trials on which both responses were correct and PIs for trials on which only one response was invalid were computed for the sample of $N = 32$ participants (see Section Data Analysis, above, for a description of the remaining sample) as a function of compatibility, task, and block-cluster. The group means of the aggregated data are shown in Table 8.

Table 8. Mean Median Reaction Times (RT, in ms) and % Invalid (PI) for Verbal (Primary) and Manual (Secondary) Responses as a Function of Response-Response Compatibility and Block Cluster in Experiment 2.

		Block Cluster									
		1		2		3		4		overall	
Response	R1-R2	RT	PI	RT	PI	RT	PI	RT	PI	RT	PI
verbal	compatible	688	4.0	650	2.3	637	1.2	635	1.8	653	2.3
	incompatible	724	6.3	688	3.4	685	3.6	668	3.1	691	4.1
	Δ	36	2.3	38	1.1	48	2.4	33	1.3	38	1.8
manual	compatible	1124	10.6	1057	4.1	1016	3.5	997	2.9	1048	5.3
	incompatible	1172	11.3	1134	5.5	1102	6.6	1075	6.3	1121	7.4
	Δ	48	0.7	77	1.4	86	3.1	78	3.4	73	2.1

Note: Rows labeled Δ indicate effect sizes of the compatibility effects. The two rightmost columns contain the overall means across block means.

RT. The median RTs were submitted to an ANOVA with task (verbal, manual) and response compatibility (compatible, incompatible), and block-cluster (1-4) as within-subjects factors. This analysis yielded a significant main effect of task, $F(1,31) = 294.74$, $p < .001$, $MSe = 73{,}994.27$, indicating faster responses on the verbal task. The main effect of block-cluster was also significant, $F(3,93) = 10.74$, $p < .001$, $MSe = 15{,}450.31$, as was the interaction between task and block-cluster, $F(3,93) = 4.75$, $p < .05$, $MSe = 4{,}648.41$, again implying that, although both tasks benefited from practice, the reduction of RT with practice was more pronounced for the manual task. As in the analysis without block-cluster as a factor (see main results), the main effect of compatibility, as well as the interaction of task and compatibility, was significant ($F(1,31) = 36.45$, $p < .001$, $MSe = 10{,}801.08$; and $F(1,31) = 22.18$, $p < .001$, $MSe = 1{,}664.95$, for the compatibility main effect and the interaction, respectively).

More importantly, block-cluster again did not interact with compatibility. That is, neither the interaction of block-cluster and compatibility, $F(3,93) < 1$, $MSe = 4{,}795.41$, nor the three-way interaction of block-cluster, compatibility, and task, $F(3,93) = 1.9$, $p > .14$, $MSe = 932.45$, reached significance, indicating that practice, while speeding up overall RTs, did not affect the compatibility effects. The latter conclusion was confirmed by separate ANOVAs on each task that did not show reliable interactions of block-cluster and compatibility (both p's > .28).

PI. As is evident from Table 8, PIs tended to follow the RT pattern. This was confirmed by the PI ANOVA, which yielded significant main effects of task, $F(1,31) = 17.97$, $p < .001$, $MSe = 71.21$, block-cluster, $F(3,93) = 36.22$, $p < .001$, $MSe = 16.8$, and compatibility,

$F(1,31) = 13.87$, $p < .001$, $MSe = 35.66$, as well as a significant interaction between task and block cluster, $F(3,93) = 4.45$, $p < .05$, $MSe = 24.03$, that signals a more pronounced reduction in errors with practice for manual responses.

Neither the task x compatibility interaction, $F(1,31) < 1$, $MSe = 25.24$, nor the interactions including compatibility and block-cluster reached significance ($F(3,93) = 1.22$, $p > .3$, $MSe = 14.44$; and $F(3,93) = 1.03$, $p > .37$, $MSe = 17.19$, for the two-way and the three-way interaction including task, respectively).

4.2.3 Discussion

When color was the relevant response dimension for the verbal and the manual task, there were again forward and backward compatibility effects of remarkable size, implying that color codes were used to control responses on the two tasks. This demonstration of color based congruency effects for manual left and right keypresses that had not been defined with respect to color prior to instruction extends studies on response-effect compatibility (see Chapter 3.1.3) that demonstrated arbitrary response coding after consistent effect-stimulus presentation.

Moreover, the IRI quintile analyses again showed that the backward compatibility effect was independent of response grouping, but rather seemed to result from parallel response activation on the two tasks that shared a common response dimension. Furthermore, neither the forward nor the backward effect was affected by practice, although overall RT level decreased with practice, again indicating that (relatively limited) practice does not lead to changes in the codes responsible for the effects. As was the case in Experiment 1, speed of responding did influence the effects. That is, the compatibility effects increased with increasing RT level, adding support to the view that enhanced processing overlap between the two tasks leads to increased crosstalk between the tasks.

Unlike Experiment 1, however, the compatibility effect in Experiment 2 significantly increased from verbal (primary) to manual (secondary) responses. Possibly, meaning-based R1-R2 priming across response modes due to a less-than-perfect reset of R1 (cf. Logan & Gordon, 2001) contributed to the compatibility effect on the manual task over and above parallel code activation (see Chapter 4.1.3, for a discussion of the lack of this interaction in Experiment 1).

In sum, the color-based backward compatibility effect suggests that instructed color codes were used to access and guide manual responses. This result extends findings on response-effect compatibility demonstrating that arbitrary action effects become included into response representations with practice. Moreover, because the instructed color codes do not contain any spatial information, the compatibility effects in Experiment 2 provide stronger evidence for an impact of response instructions on behavioral control (i.e., the direct-coding hypothesis) than the dual-task experiments that used left/right response instructions.

One argument that might be raised against this conclusion is that color label-to-key assignment varied unsystematically from trial to trial, thus giving "instructed" color coding an unfair advantage. Varied color-to-key assignment does not appear to be a necessary condition for the effects to occur, however, given preliminary results from twenty-three participants in a current replication experiment with constant color-to-key assignment. More specifically, in this replication experiment, subjects saw the same color plate (i.e., the same color-to-key assignment) before each trial, all other things being equal to the Experiment 2 described above. The results obtained so far indicate that substantial backward as well as forward compatibility effects also emerge with constant color-to-key assignment.

Another argument that can be raised against an interpretation of Experiment 2 as supporting the direct coding hypothesis is that, as in the Experiments by Logan and colleagues, as well as in many studies involving color-label responses (cf. Chapter 3), spatial response coding of manual responses has not been assessed. That is, Experiment 2 cannot decide whether color instructions can override spatial response coding. Experiment 3 addresses this issue.

4.3 Experiment 3

The goal of Experiment 3 was to assess the status of spatial response coding of manual responses under non-spatial response instructions. More specifically, Experiment 3 is concerned with the question of whether CTC effects depend on overlap of <u>instructed</u> response dimensions, and hence, whether non-spatial response instructions can reduce spatial effects or not.

To address these questions, Experiment 3 again required left and right keypresses that were instructed as blue vs. green keypresses on the manual task. This time however, the verbal task required "left" vs. "right" responses, hence spatial coding. Consequently, responses on the two tasks still overlapped regarding "location" (either responses being left or right), but no longer overlapped with respect to <u>instructed</u> response dimensions (color vs. location).

Both the spatial coding hypothesis and the weak version of the direct coding hypothesis predict a location-based congruency effect because they assume that instructions cannot override spatial response coding. If, however, response codes can be weighed according to instructions, as implied by the strong version of the direct coding hypothesis, then we should expect reduced (location based) R-R compatibility effects in Experiment 3.

Aside from response instructions, Experiment 3 was identical to Experiment 2.

4.3.1 Method

Participants

Overall, fifty-one undergraduate students (41 female, 10 male, mean age = 23.1 years) at Humboldt University, Berlin, received either € 7,- or partial course credit for participation (see Section Data Analysis, below, for a description of the final sample).

Apparatus and Stimuli

Apparatus and stimuli, as well as response keys and response recording were the same as those used in the previous experiments.

Design and Procedure

The procedure was identical to Experiment 2 with the following exceptions. The verbal task required "left" vs. "right" responses, whereas left vs. right keys were again instructed as *blue* vs. *green* keys on the manual task. Accordingly, participants first responded verbally to tone stimuli by saying either "links" or "rechts," and then pressed either the blue or the green key in response to form stimuli. Congruency was defined as in Experiment 1, that is, according to overlap regarding "location." Thus, for example, when the blue key was the one located on the left side, then a verbal "left" response followed by a blue (left) keypress, and a verbal "right" response followed by a green (right) keypress were considered congruent.

4.3.2 Results

Data Analysis

Main Results and Additional Analyses. According to pre-experimentally defined criteria, data from seven participants were excluded from Experiment 3 because they performed the manual task before the verbal task on more than 10% of the trials, and further eight participants were excluded because they produced more than 30% errors, misses and uncodable vocal responses overall for the same reasons as in the previous experiments. Statistical analyses on the data of these participants yielded similar results as the analogous analyses of included participants (see below). Two further participants had to be excluded due to experimenter error, leaving data from $N = 34$ participants (26 female, 8 male, mean age = 23.2 years) in the main analysis and additional analyses of Experiment 3.

In these data, response order errors occurred on 1.4% of the trials. These trials were excluded, as were responses that were faster than 50 ms or slower than 2000 ms (2.4%), and trials with double errors (1.2%). For the remaining data, I calculated PIs and median correct RTs for each participant as in the previous experiment.

Practice Analyses. Two further participants had to be excluded according to the exclusion criteria when data from the practice block were included. One of them now produced more than 10% response order errors overall, whereas the other participant produced more than 30% errors, misses and uncodable vocal responses across blocks.

Therefore, as in the previous experiments, data from $N = 32$ participants (25 female, 7 male, mean age = 23.4 years) were considered in the practice analyses.

In the remaining data, response order errors occurred on 1.7% of the trials and were excluded, as were double errors (1.8%) and responses that were faster than 50 ms or slower than 2000 ms (2.7%). Median RTs for trials on which both responses were correct and PIs for trials with only one incorrect response were again computed for each participant according to task, block-cluster, and compatibility.

Main Results

Table 9 presents the group means of the individual median RTs and PIs as a function of task (response) and compatibility. As is evident from Table 9, the compatibility effects on both tasks were extremely small.

Table 9. Mean Median Reaction Times (RT, in ms) and % Invalid (PI) for Verbal (Primary) and Manual (Secondary) Responses as a Function of Response-Response Compatibility in Experiment 3.

	Compatible		Incompatible		Δ	
Response	RT	PI	RT	PI	RT	PI
Verbal	656	3.4	660	3.6	4	0.2
Manual	1143	6.5	1141	5.4	-2	-1.1

Note: Columns labeled Δ indicate effect sizes of the compatibility effects (incompatible minus compatible), whereby negative values signal faster responses or fewer errors in the incompatible condition.

RT. The RT ANOVA including task as a factor only yielded a significant main effect of task, $F(1,33) = 322.97$, $p < .001$, $MSe = 24{,}731.89$. Neither the main effect of compatibility, $F(1,33) < 1$, $MSe = 811.75$, nor the interaction between task and compatibility, $F(1,33) = 1.07$, $p > .3$, $MSe = 324.42$, reached significance. Nor did the compatibility effect reach significance for either task alone, as reflected by the results of planned comparisons, $F(1,33) < 1$, $MSe = 1{,}639.51$ for the manual task, and $F(1,33) = 1.02$, $MSe = 632.82$ for the verbal task.

PI. Similarly, the overall PI ANOVA also revealed a significant effect of task, $F(1,33) = 11.08$, $p < .01$, $MSe = 18.91$, whereas the main effect of compatibility, $F(1,33) = 1.28$, $p > .25$, $MSe = 5.61$, and the interaction between task and compatibility, $F(1,33) = 2.78$, $p > .10$, $MSe = 5.06$, did not reach significance. Nevertheless, planned comparisons showed that the compatibility effect for the verbal task was not significant, $F(1,33) < 1$, $MSe = 9.02$, whereas the slightly reversed effect in manual responses approached significance, $F(1,33) = 3.36$, $p < .08$, $MSe = 12.32$.

To assess the possibility that a speed-accuracy tradeoff might have masked the compatibility effects, I correlated the RTs and PIs for each participant, task, and compatibility condition. Whereas none of the correlations between RT and PI within each task x compatibility condition reached significance (all p's > .2), both correlations for the verbal task were negative. Therefore, I also correlated the compatibility effects (i.e., the Δs) in RTs and PIs of each participant and task. The latter correlations revealed a significant positive relation for manual responses, $r = .42$, $p < .05$, whereas the relation for the verbal task was negative and approached significance, $r = -.31$, $p < .08$, suggesting that participants with large verbal RT effects tended to show small (or even reversed) PI effects and vice versa (see also practice analyses, below). However, doubly multivariate analyses of variance (MANOVAs) that simultaneously considered PI and RT as dependent variables led to similar results as the ANOVAs. That is, the compatibility effects did not reach significance, neither in the analysis including task as a factor, $F(2,32) < 1$, and $F(2,32) = 1.47$, $p > .24$, for the main effect of compatibility and the interaction of task and compatibility, respectively, nor in the analyses by task, $F(2,32) = 1.76$, $p > .18$, and $F(2,32) < 1$, for the compatibility effect in manual and verbal responses, respectively.

Comparisons between Experiments

Despite procedural differences between the experiments reported in Chapter 4 (especially between Experiment 1, on the one hand, and Experiments 2 and 3, on the other hand), additional analyses comparing Experiment 3 with Experiments 1 and 2, respectively, were run in order to substantiate the claim that response instructions reduced the compatibility effect in Experiment 3.

The 2 (experiment) x 2 (task) x 2 (compatibility) mixed-factors ANOVA[12] comparing RTs in Experiment 3 and Experiment 1 yielded significant main effects of experiment, $F(1,66) = 10.88$, $p < .01$, $MSe = 108{,}624.43$, and task, $F(1,66) = 641.64$, $p < .001$, $MSe = 19{,}890.06$, as well as a significant interaction between task and experiment, $F(1,66) = 9.06$, $p < .01$, indicating slower responses in Experiment 3, especially on the manual task.

More importantly, the main effect of compatibility, $F(1,66) = 6.49$, $p < .05$, $MSe = 6{,}131.80$, was qualified by the two-way interaction between experiment and compatibility, $F(1,66) = 5.87$, $p < .05$, whereas the interaction between task and compatibility, and the three-way interaction of experiment, task, and compatibility did not reach significance (both *F*'s < 1), implying that the compatibility effects on both tasks were reduced in Experiment 3. The PI ANOVA that yielded the same significances as the RT analysis corroborates these findings.

The comparison of Experiment 3 and Experiment 2 shows a similar picture as the comparison of Experiment 1 and 3. The RT ANOVA reveals that Experiments 2 and 3 did not differ regarding overall RT level (main effect of experiment, $F(1,66) < 1$, $MSe = 108{,}525.54$), whereas all other factors were highly significant: The main effects of task, $F(1,66) = 613.63$, $p < .001$, $MSe = 21{,}786.41$, compatibility, $F(1,66) = 21.46$, $p < .001$, $MSe = 2{,}527.39$, as well as the interactions between task and experiment, $F(1,66) = 5.32$, $p < .05$, task and compatibility, $F(1,66) = 9.72$, $p < .01$, $MSe = 455.07$, and between compatibility and experiment, $F(1,66) = 19.72$, $p < .001$, that were qualified by the three-way interaction of experiment, task, and compatibility, $F(1,66) = 18.96$, $p < .001$, implying that the differences in compatibility effects were larger for manual responses. However, when each task was analyzed sepa-

[12] Running a MANOVA instead of separate ANOVAs on RT and PI leads to the same conclusions. This also applies to the comparison between Experiments 3 and 2.

rately, the interaction between compatibility and experiment reached significance for both tasks, $F(1,66) = 10.45$, $p < .01$, $MSe = 1{,}626.83$ for verbal responses, and $F(1,66) = 23.04$, $p < .001$, $MSe = 4{,}338.09$ for manual responses, indicating that both the verbal and the manual compatibility effects were reduced in Experiment 3. The ANOVA on PI yielded similar results as the RT analysis, most notably a highly significant compatibility x experiment interaction, $F(1,66) = 16.86$, $p < .001$, $MSe = 7.3$, that was not qualified by a three-way interaction, however, $F(1,66) = 2.54$, $p > .11$, $MSe = 5.81$.

Additional Analyses

One might still argue that the fact that responses (especially manual responses) were much slower in Experiment 3, resulting in larger IRIs (on average, mean median IRI was 469 ms and 458 ms on compatible and incompatible trials, respectively; in comparison, the corresponding IRIs in Experiment 2 were 372 ms vs. 399 ms, and 354 vs. 359 in Experiment 1), led to reduced compatibility effects. Therefore, I again determined the IRI quintiles and the median RTs for each participant, quintile, task, and compatibility condition. Figure 7 shows the means of the median RTs. As becomes clear from inspection of Figure 7, little happens across IRIs except for overall slowing of manual responses. This is reflected in the results from separate ANOVAs on each task.

In the analysis of verbal responses, no effect reached significance, neither the main effects of quintile, $F(4,132) = 1.09$, $MSe = 5{,}864.69$, and compatibility, $F(1,33) = 1.6$, $p > .2$, $MSe = 1{,}730.23$, nor the interaction between quintile and compatibility, $F(4,132) = 1$, $MSe = 2{,}466.16$. In the analysis of manual responses, the effect of quintile was significant, $F(4,132) = 191.19$, $p < .001$, $MSe = 9{,}011.28$; the main effect of compatibility, $F(1,33)$, $= 1.35$, $p > .25$, $MSe = 3{,}812.92$, and the interaction between quintile and compatibility, $F(4,132) = 1.32$, $p > .27$, $MSe = 3{,}181.16$, were not.

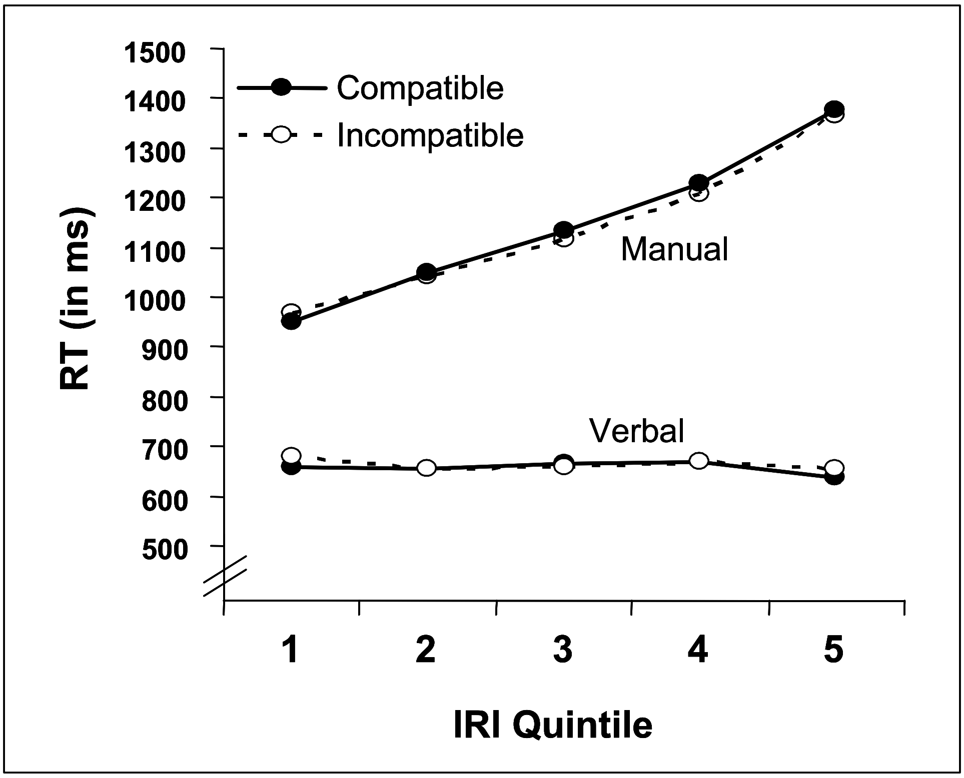

Figure 7. Mean median verbal and manual RTs as a function of IRI and compatibility between verbal and manual responses in Experiment 3. Mean median IRI quintiles were 289, 398, 469, 554, and 699 ms for compatible trials, and 294, 388, 458, 537, and 687 ms for incompatible trials.

Another objection that could be raised might be that the lack of effects may be due to trials on which the stimulus-to-(left/right) key mapping changed as a consequence of changing the color-to-key assignment on successive trials, thus requiring a re-binding of stimulus codes and left/right codes. To address this possibility, I determined median RTs and PIs according to task, compatibility, and change vs. no-change of color-to-key assignment on consecutive trials, and compared the compatibility effects for change and no-change trials. The compatibility effects did not differ across trial types. The Δs for the verbal task were 4 ms / 0.1% invalid and 7 ms / -0.1% for change and no-change trials, respectively. The corresponding Δs for the manual task were -4 ms / -2.1% and 1 ms / -0.4%. Not surprisingly, trial type did not interact with compatibility (neither in the analysis of RTs, nor in the error analysis) for either task, indicating that the null-effects were not restricted to change trials. Moreover, a similar analysis for the Experiment 2 data revealed that the color compatibility effects were not affected by the change of color-to-key assignment either.

Finally, as in the previous experiments, I ran distribution analyses to check whether the (null-) effects varied as a function of response speed (see Figure 8 for means of medians for each RT quintile, task, and condition).

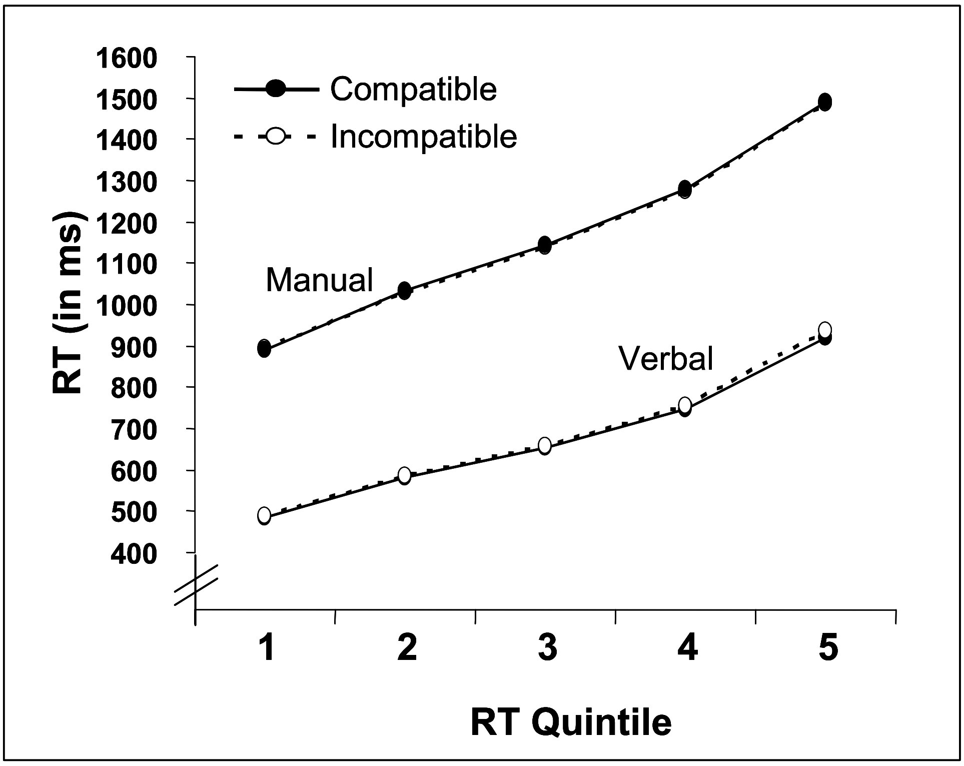

Figure 8. Mean median verbal and manual RT quintiles as a function of compatibility between verbal (primary) and manual (secondary) responses in Experiment 3.

The omnibus ANOVA including RT quintile, task, and compatibility as factors only yielded significant main effects of task ($F(1,33) = 317.4$, $p < .001$, $MSe = 125{,}967.81$) and RT quintile ($F(4,132) = 401.03$, $p < .001$, $MSe = 13{,}460.72$). It neither showed a main effect of compatibility ($F(1,33) < 1$, $MSe = 2{,}006.64$), nor did compatibility interact with task ($F(1,33) = 1.63$, $p > .21$, $MSe = 2{,}123.18$). Most important, compatibility did not interact with quintile, (F's < 1 for both, the two-way interaction of compatibility and quintile and the three-way interaction involving task). Because there seemed to be a hint of an effect at the slowest verbal quintiles, I tested whether the interaction was significant when each task was considered in isolation. These analyses showed that the interaction of compatibility and quintile did not reach significance for either task alone, $F(4,132) = 1.26$, $p > .29$, $MSe = 653.55$; and $F(4,132) < 1$, $MSe = 932.56$, for verbal and manual responses, respectively. The distribution analysis thus implies that the null effects in the main analysis were not due to a reversal of positive effects at particular RT bins.

Practice

As in the previous experiments, RTs and PIs for all blocks (including the practice block) were calculated as a function of task, compatibility, and block-cluster in order to assess how the effects develop with practice (for group means, see Table 10).

Table 10. Mean Median Reaction Times (RT, in ms) and % Invalid (PI) for Verbal (Primary) and Manual (Secondary) Responses as a Function of Response-Response Compatibility and Block Cluster in Experiment 3.

		Block Cluster									
		1		2		3		4		overall	
Response	R1-R2	RT	PI	RT	PI	RT	PI	RT	PI	RT	PI
Verbal	compatible	698	8.1	661	4.0	662	2.4	632	2.1	663	4.2
	incompatible	720	5.3	663	4.8	656	3.9	639	1.6	670	3.9
	Δ	22	-2.8	2	0.8	-6	1.5	7	-0.5	7	-0.3
Manual	compatible	1207	11.3	1165	6.4	1147	6.1	1100	5.6	1155	7.3
	incompatible	1218	11.8	1165	4.7	1130	4.2	1083	5.0	1149	6.4
	Δ	11	0.5	0	-1.7	-17	-1.9	-17	-0.6	-6	-0.9

Note: Rows labeled Δ indicate effect sizes of the compatibility effects, whereby negative values signal faster responses or fewer errors in the incompatible condition. The two rightmost columns contain the means across blocks.

RT. From Table 10, it seems as if there were compatibility effects in RTs, especially for verbal responses, on the first block-cluster. However, this impression did not receive support from the RT analysis. Although the omnibus ANOVA revealed a significant interaction of task and compatibility ($F(1,31) = 4.29$, $p < .05$, $MSe = 1{,}122.94$), in addition to significant main effects of task ($F(1,31) = 339$, $p < .001$, $MSe = 89{,}079.42$), block-cluster ($F(3,93) = 10.76$, $p < .001$, $MSe = 19{,}268.46$), and an almost significant interaction between task and block-cluster ($F(3,93) = 3.06$, $p < .06$, $MSe = 5{,}337.42$), both the two-way interaction of block-cluster and compatibility ($F(3,93) < 1$, $MSe = 5{,}067.91$), and the three-way interaction of block-cluster, compatibility, and task ($F(3,93) < 1$, $MSe = 1{,}183.62$) were far from significance. Moreover, in separate ANOVAs for each task, compatibility did not reach significance, neither for verbal responses, $F(1,31) = 1{,}53$, $p > .22$, $MSe = 1{,}830.39$, nor for manual responses, $F(1,31) < 1$, $MSe = 3{,}502.65$. Moreover, compatibility did not interact with block-cluster when each task was considered separately (both F's ≤ 1), indicating that the task x compatibility interaction in the omnibus analysis was due to nonsignificant opposite effects in (overall) verbal and manual RTs.

PI. The PI effects tended to show a somewhat different pattern than the RT effects across block-clusters, at least where verbal PIs are concerned. Moreover, the values in Table 10 indicate that the pattern of the PI effects differ across tasks. More specifically, whereas the verbal PI effects tended to show an inverse quadratic trend across block-clusters, the trend for the manual PI effect was in the opposite direction. This observation was supported by the omnibus PI ANOVA that revealed a significant three-way interaction between task, block-cluster, and compatibility, $F(3,93) = 5.66$, $p < .01$, $MSe = 12.61$, in addition to significant, but theoretically less interesting, main effects of task, $F(1,31) = 15.12$, $p < .001$, $MSe = 69.64$, block-cluster, $F(3,93) = 28.53$, $p < .001$, $MSe = 28.58$, and a nearly significant interaction between task and block-cluster ($F(3,93) = 2.8$, $p < .07$, $MSe = 30.99$; no other effects reached significance). Separate analyses for each task showed that the interaction between compatibility and block-cluster was significant for verbal responses, $F(3,93) = 5$, $p < .01$, $MSe = 11.45$. Contrasts that tested the verbal PI compatibility effects for each block (without adjusting degrees of freedom) showed that the (reversed) compatibility effect only reached significance in the first and third block-cluster, but not in the second and fourth block-cluster. In contrast, the interaction between compatibility and block-cluster was not significant for manual responses ($F(3,93) = 1.6$, $p > .2$, $MSe = 12.25$), whereas the main effect of compatibility almost was, $F(1,31) = 2.93$, $p < .1$, $MSe = 18.78$, corroborating the main results that showed a marginally significant reversed compatibility effect for manual PIs.

Because of the obvious discrepancies between RTs and PIs, and because there already was some evidence for a trade-off of effects (i.e., a slightly negative correlation between the RT effect and the PI effect on the verbal task) in the analysis without block-cluster (see Main Results, above), I again ran doubly multivariate MANOVAs that simultaneously considered PI and RT as dependent variables, this time including block-cluster as a factor. The centroids on which the MANOVA results were based closely followed the pattern of PIs because PI was weighted more strongly than RT on the discriminant function that primarily discriminated between conditions[13]. As a consequence, the omnibus MANOVA including task, block-cluster, and compatibility as factors led to similar outcomes as the PI analyses. Specifically, the interaction of task, block-cluster, and compatibility was marginally significant, $F(6,26) = 2.35$, $p < .07$. Moreover, whereas the analysis of verbal responses again revealed a

[13] MANOVAs on the practice results of Experiments 1 and 2 that were run for reasons of comparability also revealed a relatively strong dependence on PI. However, in those experiments PI and RT effects go in the same direction. Therefore, those MANOVA outcomes do not contradict the RT analyses.

significant interaction of block and compatibility, $F(6,26) = 3.6$, $p < .01$, the interaction did not reach significance for manual responses, $F(6,26) < 1$.

4.3.3 Discussion

In the RT ANOVAs neither the 4 ms backward compatibility effect on the verbal task, nor the –2 ms forward compatibility effect on the manual task reached significance, and the effects were significantly smaller than in Experiments 1 and 2. The MANOVAs that conjointly considered RT and PI suggest that the lack of effects in Experiment 3 cannot be attributed to a tradeoff between (effects in) RTs and PI. In addition, the significant experiment x compatibility interactions in the experimental comparisons, combined with the facts that (a) the *N*'s were rather large, and that (b) the statistical error terms (i.e., the *MSe*'s) for the compatibility effects in Experiment 3 did not exceed those of the previous experiments, indicate that the null effects in Experiment 3 were not due to power problems. Finally, the lack of effects was not restricted to trials on which the color-to-key assignment, and hence the stimulus-to-location mapping, changed, but also held for no-change trials, indicating that re-binding of location codes to different stimulus attributes cannot be the main cause for the outcome.

Rather, the results seem to suggest that R-R compatibility effects are extremely reduced when instructed response dimensions do not overlap, even when the two tasks share a common implicit response dimension (i.e., location). This result corroborates the findings by Logan and colleagues (e.g., Logan & Schulkind, 2000) who also demonstrated a lack of intertask effects when categorization tasks changed from Task 1 to Task 2.

Interestingly, regarding the forward compatibility effect (i.e., the effect in manual responses) there even was a (marginally significant) tendency for a reversal of the compatibility effect in PIs that also manifested itself in the mean median IRIs (remember that mean median IRI was 469 ms and 458 ms on compatible and incompatible trials, respectively). This observation suggests that manual responding was slightly more difficult on compatible than on incompatible trials, and may imply that some sort of reset mechanism inhibits response "repetition" on compatible trials. I will return to this issue in the General Discussion section (see Section 4.4).

Unlike Experiments 1 and 2, the compatibility effect in PIs on the verbal task seemed to inconsistently change with practice, namely from negative to positive back to negative. However, this pattern did not show up in RTs and was inconsistent across tasks. At present, it is unclear, which factors (e.g., strategies) may have led to this pattern of results, and whether it is systematic or not.

Nevertheless, whatever the exact reasons for the practice results, the IRI quintile analyses suggest that the lack of RT effects (or their reduction) was not due to the fact that verbal and manual responses in the main analysis were, on average, scheduled further apart in Experiment 3 than in the other two experiments, thus resulting in reduced overlap in Task 1 and Task 2 processing. First, the interaction between IRI quintile and compatibility was not significant for either task. Second, although there seemed to be a hint of a compatibility effect at the first IRI quintile (the mean median IRI of which was 291 ms), in the other two experiments the compatibility effects were significant at much higher IRIs (e.g., both the verbal and the manual compatibility effects reached significance at the highest IRI quintile, the mean medians were 586 ms and 596 ms in Experiments 1 and 2, respectively). Similarly, the distribution analysis indicates that the size or the direction of the null-effects did not significantly differ as a function of response speed.

In sum, instructing spatially organized responses in terms of color reduces the spatial (backward and forward) compatibility effects observed under spatial response instructions, suggesting that manual responses were arbitrarily coded as instructed, thereby providing strong support for the intentional weighing hypothesis.

4.4 General Discussion Experiments 1-3

Experiments 1-3 used a dual task approach requiring consistent viz. inconsistent responses on two tasks to investigate which of three different views on how response labels used in task instructions influence response coding is correct. According to the spatial coding hypothesis, responses are spatially coded, regardless of instructions. Therefore, the spatial coding hypothesis predicted cross-task compatibility effects on Experiment 1 and 3 where responses overlapped in terms of 'location'. In contrast, no congruency effects were expected with respect to arbitrary instructed response dimensions such as color (Experiment 2). The direct coding hypothesis, on the other hand, assumes that arbitrary response codes are included into response representations when so instructed. Hence, cross-task compatibility should extend to arbitrary response dimensions (Experiment 2). Whereas the weak version of the direct coding hypothesis assumes that non-instructed (i.e. spatial) response codes are weighed as strongly as explicitly mentioned response dimensions, the strong version proposed that response codes can be intentionally weighed. Accordingly, inter-task consistency effects resulting from implicit (non-instructed) overlap on the spatial dimension (Experiment 3) were expected by the weak direct coding hypothesis, but not by its strong version.

In Experiment 1, both a verbal and a concurrently performed manual task required left and right responses to univalent, non-integrated stimuli that were arbitrarily mapped to responses. Responses on both tasks were faster and less error prone when the two responses were left or right as opposed to one being right and one being left, thus replicating Hommel (1998, Exp. 1) using a slightly modified paradigm. Experiment 2 showed that such (forward and backward) R-R compatibility effects are not restricted to the spatial response dimension, but also occur with abstract response dimensions that cannot be assumed to have been part of the manual response representations prior to instruction (i.e., color). This result extends findings on response-effect compatibility showing that arbitrary attributes become integrated into response representations after training. Both forward and backward compatibility effects were extremely reduced when different response dimensions were instructed for the two tasks in Experiment 3 (i.e., when the verbal task required "left" and "right" responses, whereas the left and right keys on the manual task were instructed as blue vs. green). The latter finding corroborates the results obtained by Logan and colleagues and suggests that inter-task consistency effects depend on overlap of instructed response dimensions.

<u>Implications for Response Coding</u>. The compatibility effects in Experiments 1 and 2 suggest that the same relatively abstract conceptual response codes (abstract in the sense that they do not necessarily contain information necessary for motor responding) were used for response selection on the verbal and the manual task. Furthermore, they suggest that the same responses can be controlled differently, depending on response instructions, indicating a high degree of flexibility in response coding. In line with the prediction of the direct coding hypothesis the color dimension was used for response selection on the manual task when response instructions primed the color dimension for both tasks in Experiment 2.

If one adopts the view that responses are represented in a distributed fashion (Allport, 1993; Hommel et al. 2001; Keele, Cohen, & Ivry, 1990) such that every response is coded in terms of its features (e.g., as being blue, left, manual, …), then the compatibility effects in Experiment 1 and 2 can be explained in a similar way as SRC effects, namely as a result of a conceptual match (vs. mismatch) of overlapping response features that are part of the response representations of both tasks (cf. Lien & Proctor, 2002, for a similar view). According to this view, response representations that share features across tasks tend to be activated in parallel by the stimulus representations that are assigned and temporarily linked to them, leading to facilitation if the same code is activated by both stimuli, but to response competition if not (see Figure 9, panel A and B for an illustration of how this might lead to response competition in the incompatible conditions of Experiments 1 and 2). Note that this view, in accordance with most coding accounts of compatibility, implies that response selection primarily occurs at the level of conceptual response codes that (automatically) activate their 'corresponding' motor programs (e.g., the left and right hand motor codes Mm_l and Mm_r in Figure 9, Panel A and B), rather than at the level of motor programs.

Under this assumption, the lack (or the reduction) of the spatial backward compatibility effect in Experiment 3 suggests that arbitrary response codes were not only part of the manual response representation under color instructions, but that color coding actually dominated spatial coding. That is, in line with the predictions of the strong version of the direct coding hypothesis, Experiment 3 showed that R-R compatibility effects are only observed when <u>instructed</u> response dimensions overlap, indicating that (a) manual responses were primarily coded and accessed in terms of color even when the verbal task required spatial coding, and (b) location codes did not play an important role in selecting manual responses when responses were instructed in terms of color. Figure 9 (Panel C) depicts how the lack of effects in Experiment 3 might be explained. More specifically, the stimuli for the two tasks in Experiment 3 were assigned to different response dimensions. Hence, responses for the two tasks were accessed by different codes, leading to less overlap regarding activation and use of location codes.

I do not argue, however, that location codes were completely substituted by color codes and omitted from manual response representation and selection in Experiment 3 (and Experiment 2). Rather, I believe that they were still part of the manual response representation – though less strongly weighed –, and were integrated with color codes (see Figure 9, Panel B and C) to allow manual responding, much as anatomical codes contribute to the Simon effect under crossed-hand conditions (see Chapter 3.1.2). This is so because if no spatial codes were included in the manual response representation at all, the response representation of the two tasks would not share codes. Accordingly, one would expect a clear-cut null-effect in both verbal and manual responses. However, in Experiment 3 there was a tendency for a reversed compatibility effect, that is, costs on compatible as compared to incompatible trials, in manual task errors and in mean median IRIs.

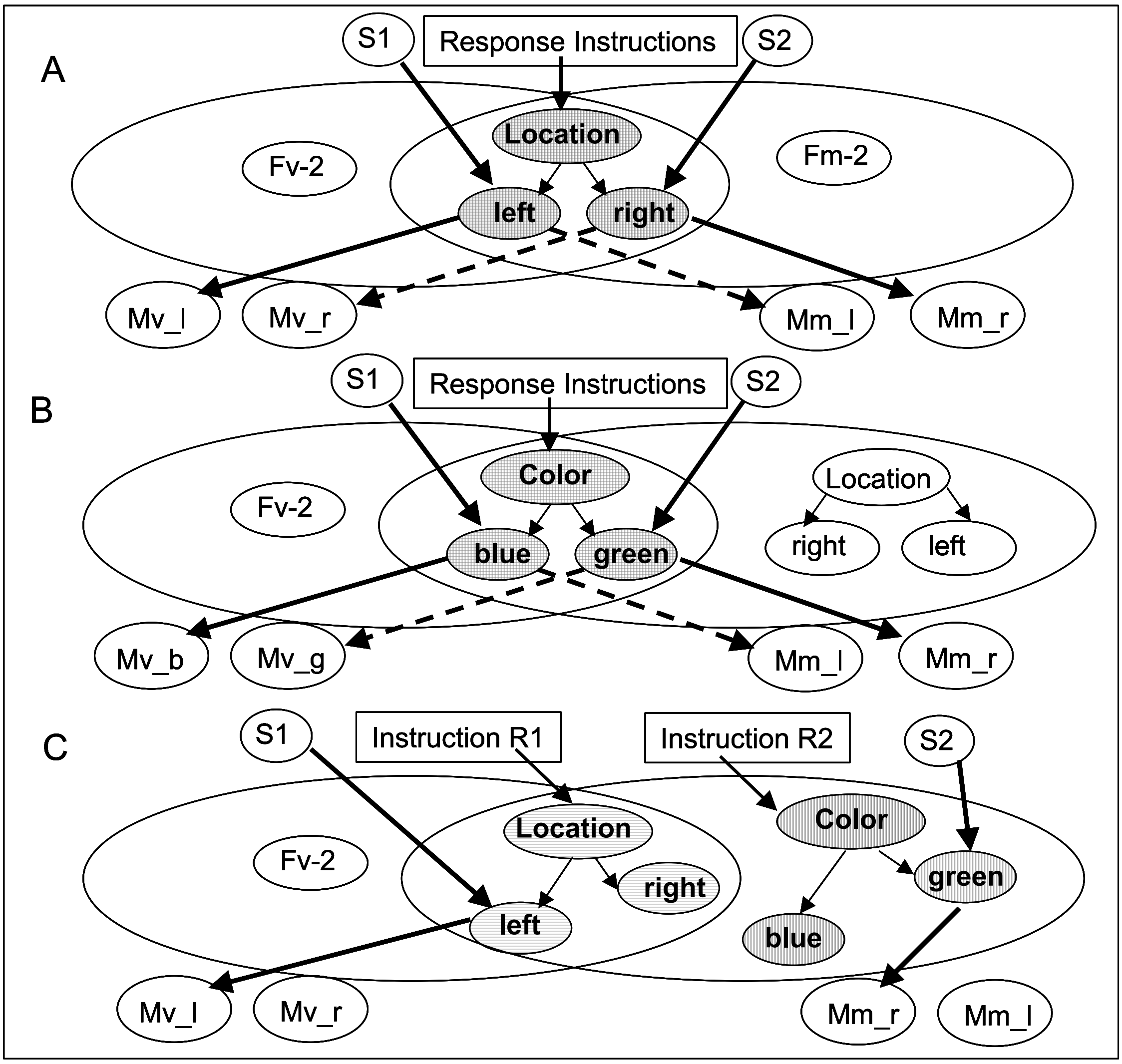

Figure 9. An illustration of the hypothetical activation flow in incompatible conditions in Experiment 1 (Panel A), Experiment 2 (Panel B) and Experiment 3 (Panel C). Panel A: In Experiment 1, both tone stimuli (S1) and form stimuli (S2) activate highly weighted (conceptual) location codes, which in turn activate their corresponding motor programs (Mv and Mm for verbal and manual responses, respectively; subscripts _l and _r, stand for left and right, respectively). On incompatible trials, different location codes receive activation from the stimuli, leading to interference on the two tasks. Panel B: When color is the instructed response dimension for both tasks, S1 and S2 activate the color codes that are assigned to them, leading to slower responses on both tasks if different color codes are required for responding. Panel C: Response instructions prime the location dimension for the verbal task and the color dimension for the manual task, and S1 activates a location code, whereas S2 activates a color code. Therefore, S1 and S2 do not provide diverging evidence for codes on the common location dimension (see text for details). *Note*: Fv-2 and Fm-2 represent response features that are unique to verbal and manual response representations, respectively.

Response repetition costs have also been reported in the task switching literature (cf. Meiran, 2000), where enhanced switch costs were observed when the same response had to be repeated on switch trials. Recently Schuch and Koch (submitted) observed similar response repetition and even (forward) response compatibility costs in dual task performance using a PRP procedure more similar to my experiments. Most of these studies used bivalent stimuli (e.g., numbers) on two tasks (e.g., magnitude judgments on Task 1 and parity judgments on

Task 2), and interpret response repetition or response compatibility costs as indicating re-coding of the responses in terms of what they signal (see Logan & Gordon, 2001, for a potentially similar interpretation of the lack of correspondence effects when categorizations differ across tasks). Accordingly, a left response that is recoded as, say, *odd*, on Task 1 hampers performance on a subsequent task when this task requires a left response to signal the property *small*, either because the competing (now relevant) category-response rule has been inhibited during Task 1 processing, and/or because re-binding of the left code with a different meaning during Task 2 processing requires prior unbinding of this code.

While the tendency for costs in manual responses in my Experiment 3 cannot be explained as resulting from changes in what they signal across tasks (otherwise there should have been forward compatibility costs in Experiments 1 and 2, too), unbinding and re-binding of (overlapping) location codes into a response representation that is considered “different” because of response instructions (e.g., integrating a left code into a *green* response after accessing a verbal “left” response via the left code) may have contributed to costs in my Experiment 3 as well. Alternatively, some sort of reset mechanism similar to that proposed by Milliken and colleagues (Milliken, Joordens, Merikle, & Seiffert, 1998) in the domain of negative priming might have adjusted the weights of location codes and thus hampered manual performance on compatible trials. Whatever the exact nature of the reset mechanism (see Schuch & Koch, submitted, for a comprehensive discussion of possible mechanisms underlying response repetition costs), the tendency for compatibility costs indicates that spatial codes were not completely excluded from manual response representations.

In sum, the color-based backward compatibility effect in Experiment 2 and the lack of a spatial backward compatibility effect in Experiment 3 indicate that manual responses were primarily coded in terms of color when so instructed. Location codes may still have been part of the manual action representations, but played a minor role in accessing manual responses, indicating that subjects have considerable control over how they code and access their responses. Together, Experiments 2 and 3 thus provide evidence in favor of the strong version of the direct coding hypothesis and against obligatory spatial response coding.

As already noted in the discussion section of Experiment 2, one argument that could be raised against this conclusion is that spatial coding did not have a “fair” chance in the current Experiments 2 and 3. More specifically, one may argue that unpredictably changing the color-to-key assignment from trial to trial might have reduced the likelihood of spatial response coding. While, in principle, participants could have intentionally recoded instructed mappings (e.g., square < > blue) into location based coding (e.g., square < > left) when they were informed about the color-to-key mapping on the forthcoming trial, varying the color-to-key assignment strongly discouraged such a strategy. If the pattern of results in my Experiments 2 and 3 resulted from varied color-to-key assignment one would expect no or extremely reduced color-compatibility effects and substantially enhanced spatial compatibility effects on replications of Experiments 2 and 3 with constant color-to-key assignment, respectively. That is, when both the color and the location route are viable alternatives to response selection, but instruction suggests color coding. This is what I am currently testing in replication experiments in which the color plates presented in the beginning of each trial always signal the same assignment.

As already mentioned (see Section 4.2.3 above), preliminary data from the Experiment 2 replication experiment with constant key-assignment requiring color responses on both tasks indicate substantial color-based forward and backward compatibility effects, suggesting that manual responses are coded in terms of color even when spatial re-coding of S-R mappings is easy. Moreover, while at present there are not yet any data available from the replication of Experiment 3, the experiments presented in Chapter 5 below indicate that varied assignment was not a key factor for the reduction of the effects in the current Experiment 3, either. In the experiments presented below the color-to-key assignment remained constant throughout a complete block of trials.

Implications for S-R Translation. The backward compatibility effects (i.e., the compatibility effects on the verbal task) in the first two experiments suggest that S2-R2 translation overlapped in time with S1-R1 processing. The compatibility effects seem to be larger when verbal RTs are slow, indicating that the effect increases with increased overlap in processing of the two tasks. This finding appears to be consistent with other results (e.g., Lien & Proctor, 2000; Logan & Schulkind, 2000) that show larger compatibility effects at short as opposed to long SOAs. Moreover, the compatibility effects in Experiments 1 and 2 (as well as the lack thereof in Experiment 3) were largely independent of IRI, indicating that such effects do not depend on temporal response grouping, that is, on withholding R1 execution until R2 is selected. Rather, these results imply parallel activation or retrieval of R1 and R2 information in dual-task performance, leading to compatibility effects when task relevant response dimensions overlap, and to a lack of effects when they do not overlap.

Evidence for parallel response activation is difficult to reconcile with strong response selection bottleneck interpretations of the PRP effect that assume that S2-R2 translation has to wait until S1-R1 translation is completed. I do not argue, however, that parallel S-R translation eliminates the usual PRP effect. While I did not include an SOA manipulation in the present experiments, other studies did (e.g., Lien & Proctor, 2000; Logan & Schulkind, 2000). Their results show a typical PRP pattern (i.e., slowed R2 responses when SOA is short) over and above inter-task consistency effects. Hommel (1998; see also Lien & Proctor, 2002) therefore suggested to distinguish between response activation, on the one hand, and response identification or decision processes, on the other hand, much as the DO model of SRC does (e.g., Kornblum et al., 1990; also see Logan & Gordon, 2001). According to this view, response activation proceeds in parallel, whereas response identification is serial. Consistency effects have been attributed to parallel response activation processes, whereas the PRP effect has been explained by serial response identification.

Interestingly, the size of the consistency effects in my Experiments did not (consistently) change with practice. This finding is consistent with the results obtained by Hommel (1998), and presumably implies that, whatever changes in processing (such as short cuts, instance based processing, or strengthening of rules) may lead to more efficient translation with practice (i.e. faster overall RTs), these changes do not affect the codes used for response selection in tasks like the ones used here. Furthermore, – when present at all – substantial forward as well as backward effects were already observed in the first block cluster. Therefore it seems as if automatic translation develops relatively quickly even when S-R mappings are arbitrary. This conclusion receives support from studies that demonstrated automatic response activation according to implemented mappings after relatively little practice even when (a) no responses were required on the secondary task on a complete block of trials (Azuma, Prinz, & Koch, in press), (b) the old mapping was no longer valid (i.e., when new mappings had been instructed for the secondary task; Hommel & Eglau, 2002; also see Wenke & Frensch, 2000, for analogous findings with a paradigm more similar to the one used here), and when (c) memory load was increased (Hommel & Eglau, 2002).

According to one interpretation of such findings (cf. footnote 6, Chapter 3.2), direct S-R links are established after relatively little practice that lead to automatic response activation although stimuli and responses do not (conceptually) overlap (i.e., no SRC exists; Hommel & Eglau, 2002; Proctor & Lu, 1999). Alternatively, activation is transmitted automatically and in parallel along the translation (conditionally automatic) route once it is implemented (Lien & Proctor, 2002; Tagliabue et al., 2000). Although both explanations are generally consistent with a response coding interpretation of my results, according to the latter an alternative interpretation is also possible.

More specifically, according to the 'automatic translation' account it could be argued that such effects as those observed in Experiment 2 (and the lack thereof in Experiment 3) are 'located' at some intermediate translation stage instead of reflecting type of response coding. Such a view would, for instance, attribute inter-task consistency effects to verbal codes (i.e., location or color names) that mediate responding on the manual task effects, and not necessarily to response coding per se. Although this possibility seems unlikely, given that it is only tenable if one assumes that incorrect location or color names are even retrieved when no longer valid (Wenke & Frensch, 2000) or needed (Azuma et al., in press), it cannot be ruled out with the experimental approach chosen in Experiments 1-3.

Therefore, Experiments 4 and 5 (Chapter 5) tried to extend the present findings to a task that is commonly accepted to be associated with automatic response activation, namely the Simon task.

5 One-trial-Simon Experiments

An even stronger case for a direct influence of response instructions on response coding could be made if it were possible to show that the impact of an irrelevant spatial stimulus attribute (i.e., the Simon effect) depends on response instructions. That is, if non-spatial response instructions reduced the spatial Simon effect, for the following reason.

Remember that the widely accepted dual route models (cf. Chapter 2.2) assume a direct link between overlapping stimulus and response codes. The primary motivation for dual route models to propose a direct route over and above a (controlled or conditionally automatic) translation route has been to parsimoniously account for an impact of irrelevant stimulus attributes. For example, in the standard Simon task in which the relevant S-R mapping is arbitrary, stimulus position is uninformative with respect to the required responses, and the likelihood of accidental co-translation due to logical recoding is low.

Strong behavioral evidence for the direct response activation account (and against translation accounts of the Simon effect; e.g., Wallace, 1971) comes from studies in which irrelevant spatial information is provided by go/no-go signals that are presented some time after the imperative stimulus. Hence, irrelevant spatial information is presented when imperative stimulus identification and translation can be assumed to be completed on most trials. For example, in experiments conducted by Hommel (1995, 1996c; also see Shiu & Kornblum, 1999) participants were required to press a left or a right key as indicated by spatial precues (i.e., arrows pointing to the left or right; direct mapping), but had to withhold their responses until they saw a go vs. no-go signal (i.e., a green or a red color patch; Hommel, 1995, Exp. 1), or until a go-signal was presented (i.e., a green patch; Hommel, 1996c, Exp. 1) that randomly appeared on the left or the right of the screen after varying intervals. Hommel observed a substantial Simon effect (i.e., a correspondence effect between response location and go-signal location) even when go-signals were presented on 100% of the trials. That is, when subjects presumably had a high motivation to prepare their responses in advance. Hommel concluded that irrelevant response-overlapping stimulus information automatically activates corresponding response (action) codes, thereby influencing (spatial) responding as long as the response has not been executed.

As outlined in Chapter 3, the findings regarding the influence of response instructions on the Simon effect can be considered inconclusive. On the one hand, Hommel (1993a) could show that variations of spatial response instructions determine the direction of the Simon effect. However, the Hommel experiment manipulated instructions within the spatial dimension. Simon et al. (1976), on the other hand, found a substantial spatial Simon effect when keys were instructed in terms of color and color-(key)-labels changed from trial to trial. However, the latter study did not include a (baseline) condition with spatial response instructions, leaving open the question of whether non-spatial response instructions modulate the Simon effect.

Experiment 4 and 5 extend these findings by using a variant of the Simon task similar to that of the Hommel (1995; 1996c) experiments, and by employing a response-instruction logic similar to that of Experiment 3 (see Chapter 4.3). As in the Hommel (1995) experiment, irrelevant spatial information in Experiments 4 and 5 was provided by go/no-go signals. Go/no-go signals consisted of vertical and horizontal bars presented at different locations, the orientation of which indicated whether the prepared response was to be executed or not. Go/no-go signals followed the imperative stimuli after a considerable delay. Letters served as imperative stimuli. New letters (i.e., letter pairs), and hence, new S-R pairings, were instructed on successive trials in order to avoid the usual confound between effects of instructions and practice.

Experiments 4 and 5 differ regarding response instructions while stimuli and stimulus presentation remained constant across experiments. That is, in Experiment 4, keys were instructed in terms of location. In contrast, the same keys were instructed in terms of color in Experiment 5. Experiment 4 thus establishes a baseline for a “standard” Simon effect under conditions of changing imperative stimuli. Experiment 5, on the other hand, assesses whether non-spatial response instructions affect the size of the Simon effect.

Both the spatial coding hypothesis and the weak direct coding hypothesis again predict normal and comparable spatial Simon effects in both experiments, either because instructions do not affect response coding (spatial coding hypothesis), or because non-spatial coding does not affect implicit spatial coding (weak direct coding hypothesis). Alternatively, if color coding dominates spatial coding under non-spatial response instructions (i.e., if spatial codes are less strongly weighed and hence contribute less to responding), as assumed by the strong direct coding hypothesis, the (spatial) Simon effect should be strongly reduced under non-spatial as compared to spatial response instructions.

5.1 Experiment 4

The main goal of Experiment 4 was to secure that a standard Simon effect shows up when S-R rules, and hence imperative stimuli, change from trial to trial. Responses were instructed in terms of key location, hence providing a baseline for the Simon effect under non-spatial instructions (Experiment 5). The task was similar to the one used by Hommel (1995, 1996c) in that irrelevant spatial information was provided by go/no-go signals that followed the imperative stimuli after a considerable delay. Unlike Hommel, relevant S-R mappings were arbitrary instead of compatible. That is, letter stimuli (instead of arrows) were assigned to left and right keys before each trial, thus, according to the dimensional overlap model taxonomy (cf. Kornblum & Lee, 1995), changing the Hommel task from a spatial Stroop task (type 8 ensemble) to a regular Simon task (type 3 ensemble). If the Simon effect obtained by Hommel (also see Shiu & Kornblum, 1999) was indeed due to an overlap between irrelevant stimulus position and response location, a Simon effect should also be observed with arbitrary relevant S-R mappings.

A further modification of the Hommel (1995, 1996c) task concerns the inclusion of a neutral condition. The neutral (irrelevant position) condition was realized by presenting go/no-go signals at the central screen position. It was included to assess whether potential compatibility and instruction effects obtained with this paradigm are primarily due to interference or to facilitation.

The choice of a delayed-position paradigm with changing S-R mappings was motivated by several considerations. First, imperative stimuli changed from trial to trial in order to assess the impact of response instructions without any confounding influence of practice. This should not affect direct response activation per se because the direct route is considered independent of whether new imperative stimuli have to be translated on each trial, or whether well-practiced S-R rules are retrieved. However, with new instructions on every trial, S-R translation can be assumed to be quite time consuming. This could have been problematic with a standard Simon task (i.e., a task in which imperative stimuli randomly appear at different screen positions) because it has been shown that irrelevant spatial code activation is transient and relatively short-lived. For example, Hommel (1993b) demonstrated that the Simon effect decreases as the difficulty of identifying the relevant stimulus, and hence RT, increases. Therefore, giving the imperative stimulus a head start in my experiments should enhance the likelihood to observe effects of spatial code activation.

Moreover, I employed a high percentage of go-trials and a large interval between imperative stimuli and lateralized go/no-go signals. This was done to maximize the likelihood that participants complete stimulus identification and S-R translation before the go-signal (i.e., irrelevant position information) appears. If any impact of instructions shows with this paradigm (see Exp. 5), they are likely due to influences on response coding, rather than translation.

5.1.1 Method

Participants

Overall, thirty-one subjects, students and non-students from Berlin (17 female, 14 male, mean age = 24 years), received € 7,- for participation (for exclusion and substitution of participants, see Section Data Analysis, below).

Stimulus Material and Counterbalancing

For the experimental trials, forty-eight letter pairs were constructed by pairing twenty-four letters as follows. Twelve letter pairs were generated for each of the four experimental blocks (see next section) according to several criteria. Letters in a letter pair were required to be at least 6 letters apart in the alphabet (mean distance = 11.2 letters) to avoid associative and/or ordinal (location) priming between neighboring letters. Each letter was paired with three other letters, such that, on each block, it only appeared in one letter pair.

For each block, letters in a letter pair were assigned to left or right responses such that, for half of the letter pairs, left responses were assigned to the letter that occurs "left" in the alphabet (e.g., A: left key; N: right key), whereas right responses to the "left" letter were required on the other half of the letter pairs (e.g., R: left key; C: right key). When re-pairing letters to construct pairs for successive blocks, half of the letters were assigned the same response location as in the previous block, while for the other half the assignment was changed. An attempt was made to ensure that (a) changes of assignment affected ordinally corresponding S-R assignments (i.e., A: left key; N: right key) approximately as often as non-corresponding assignments, and that (b) response assignment for a particular letter changed at least once across blocks. The latter worked out for 22 out of the 24 letters.

For each of the twelve letter pairs per block, all six go-signal-position (left, middle, right) x response location (left vs. right) conditions were generated. These six go-signal-position x response location conditions were distributed across three lists by Latin square such that each letter pair appeared only twice within a single list and block. The two repetitions of each letter pair were such that (a) each letter of a given pair was presented only once (i.e., each letter pair appeared in one left-response and one right-response condition), and (b) both instantiations of a specific letter pair realized different compatibility conditions (compatible, neutral, incompatible). The Latin squares used for counterbalancing guaranteed that (a) different combinations of compatibility conditions were determined for different letter pairs on a given list and block, and (b) the same number of trials per stimulus (go-signal) position x response location condition appeared in each block and list. This counterbalancing scheme led to 24 trials per block and list (i.e., 4 of each stimulus position x response location condition in each block, 16 across blocks), whereby, within a given list, different conditions were mostly realized by different letter pairs. Across lists, however, each letter pair appeared in all conditions.

Of the 24 trials within a given list and block, three trials were chosen to be no-go trials (i.e., imperative stimuli were accompanied by horizontal instead of vertical bars at the predetermined position), amounting to twelve (12.5%) no-go trials per list. The three no-go trials per block and list were determined such that, across blocks, the same number of trials in each stimulus position x response location served as no-go trials (i.e., two trials per condition) and that, within a given block and list, (a) different letter pairs were used for no-go trials, and (b) the excluded responses varied in type (i.e., either two left and one right responses, or one left and two right responses were excluded, never three left or three right responses). Determining no-go trials this way left 14 go-trials (87.5%) per stimulus position x response location condition within each list, that is, 28 go-trials per compatibility condition (i.e., compatible, neutral, incompatible).

One may argue that this is a relatively small number of trials per compatibility condition. The choice to go with such a small number of trials per conditions was based on two considerations. First, there appeared to be a trade-off between reliability of estimates (enhanced with increasing number of trials) and the possibility of carry-over effects if the same letters were presented too often (albeit in different letter pairings). Second, psycholinguistic studies that employ similar counterbalancing schemes as the one used here often only present between 10 and 20 instances per condition as well. Nevertheless, they tend to attain reliable results. However, in order to back up their main results, they often carry out item-analyses (with items instead of, or in addition to, participants as a random factor; e.g., Clark, 1973) to ensure that the restricted sample of selected items represents the larger population of items (in my case letter pairs) sampled from. Therefore, I decided on fewer letter repetitions, but to carry out corresponding item analyses for the present experiments (see Section Data Analysis for details).

From the two letters that were not used in constructing experimental trials (i.e., the letters Z and I), three trials were constructed (one for each compatibility condition, with one trial being a no-go trial) that served as practice trials in the experiment and were not counterbalanced across lists.

Apparatus and Procedure

The experiment was run on Pentium II computers that were connected to two separate keys via an ExKey Logic provided by BeriSoft. Stimulus presentation and response recording was controlled by the Experimental Runtime System software (ERTS©; Beringer, 2000). The ERTS software runs in DOS mode on IBM compatible computers.

Letters appeared in white against black background at the center of the screen. Letter font was the NRC7BIT large-scale bitmap font. The height of the letter stimuli on the screen was 1.2 cm, whereas letter width ranged between 0.6 cm and 1.1 cm, depending on letter identity. The diameter of the line print was about 2 mm. Vertical bars of 3.7 x 1.7 cm (height x width) size served as go-signals, whereas horizontal bars of the same size (width x length) as the vertical bars served as no-go signals. Go- and No-go-signals were also presented in white against black background, and appeared at one of three horizontal positions. The screen center served as the neutral position; the left position was 10 cm on the left of the screen center, and the right position was 10 cm to the right. Viewing distance was approximately 50 cm.

Left and right response keys were two separate external keys. Each key was mounted on a flat metal plate, which was connected to the computers via a so-called ExKey Logic, a system also provided by BeriSoft (Beringer, 2000). The response keys were located to the left and the right of a participant's body-midline and were aligned with the screen. The distance between the two keys was approximately 20 cm.

After receiving written as well as oral instructions on the task requirements and the sequence of events within a single trial that emphasized both, the need for speed and accuracy, participants worked through five blocks. The first block consisted of three trials that were the same for all subjects. This block served as a practice block that was administered to acquaint participants with the procedure (see Section Stimulus Material and Counterbalancing, above, for the construction of trials). Participants were then given four experimental blocks consisting of twenty-four trials each. Experimental trials differed across subjects as a function of the list participants were assigned to. The three lists were approximately counterbalanced across participants by Latin square, such that about eight participants out of twenty-four subjects overall saw a particular list (see Section Stimulus Material and Counterbalancing, above, for counterbalancing of trials across lists, and Section Data Analysis, below, for a description of the final sample of subjects) [14].

Each trial started with a written instruction stating the mapping of responses (left or right key) to the letters for the upcoming trial. The stimulus-response mappings for a given pair (e.g., R: left key; C: right key) were presented below each other at the center of the screen. For half the trials in each block, the left-key S-R mapping was presented above the right-key assignment, whereas it was presented below the right-key assignment on the other half of the trials. Care was taken to ensure that the instruction-order counterbalancing applied as often to order-corresponding S-R mappings (i.e., letters left and right in the alphabet assigned to left and right responses, respectively) as to noncorresponding mappings (i.e., left and right letters assigned to right and left responses, respectively).

The mapping instructions remained on the screen for at least 2 seconds. After two seconds had passed, subjects could commence the trial when they felt they had remembered the instructions by pressing either the left or the right key, depending on their counterbalancing condition. Half the subjects in each list condition (i.e., four participants on each list) pressed the left key to initiate a trial, whereas the other half pressed the right key. Pressing the key triggered a fixation cross at the central screen position that remained on the screen for 500 ms. Simultaneously with the offset of the fixation cross, one of the letters of the instructed letter pair appeared at the same position for 500 ms. Letter offset was followed by a 1200 ms blank interval after which a go-signal (vertical bar) or a no-go signal (horizontal bar) was presented for 150 ms at one of the three horizontal positions (left, middle, right) that indicated whether the (prepared) response should be executed or not. Response recording was initiated at go/no-go signal onset and was terminated when a response was made or after 950 ms had passed. If participants made the wrong response (i.e., pressed the wrong key on go-trials), or if their go-

[14] Due to experimenter error, one participant that should have received a particular list according to the counterbalancing scheme was mistakenly given another list.

responses were either too slow (i.e., slower than the 950 ms recording duration) or too fast (i.e., when a response was made before the go/no-go signal appeared) they received a written error feedback in red against black background on the lower part on the screen for 1000 ms, followed by a 500 ms blank interval after which the mapping instruction for the next letter pair (i.e., the next trial) appeared on the screen. When participants incorrectly pressed a key on no-go trials, a 1500 Hz warning tone was presented for 500 ms via the internal PC speaker, followed by a blank interval of 500 ms until the next trial started. In the case of correct (non-) responses, the next mapping instructions were presented after a blank interval of 500 ms that followed the response or the maximum recording duration.

Trials were presented in a quasi-random pre-determined order, which ensured that (a) at least six trials/letter pairs intervened between the two repetitions of a single letter pair within a block (mean number of intervening trials = 10.4 on each list), (b) trials realizing the same stimulus (go-/no-go signal) position x response location condition did not follow each other more than once in a row, and (c) no-go trials neither occurred on the first nor the last position within a block and did not follow each other immediately.

5.1.2 Results

Data Analysis

Six participants were excluded from the analyses because they produced errors or misses on more than 10% of the trials. This relatively strict exclusion criterion seemed warranted given the low number of trials overall, leading to unreliable RT estimates when participants make too many errors. One additional participant was excluded because of very slow responses (i.e., his or her average RT across conditions exceeded the overall group mean by more than 2.5 standard deviations), indicating consistent postponement of S-R translation until go-signal presentation. However, inspection of the data of excluded participants revealed a qualitatively similar pattern of results as that observed for included participants.

The raw RTs of correct go-responses of the remaining twenty-four participants (14 female, 10 male, mean age = 23.1 years) were trimmed by excluding trials that were more than 2.5 standard deviations away from a participant's overall mean across compatibility conditions. This screening procedure was applied in order to eliminate individual RT outliers that may have unduly influenced the RT estimates, given the small number of trials used in the present experiments (cf. Ratcliff, 1993, who argued that data trimming helps to stabilize RT estimates). Overall, 2.6% correct responses were screened, and the pattern across compatibility conditions resembled that of the errors (see Section 4.2.2). No-go errors (i.e., false alarms) occurred on 32% of the no-go trials, that is, on about four of the twelve no-go trials, and were not analyzed further.

For the main analysis (i.e., the subject analyses), median RTs for each participant were computed on trimmed correct raw go-RTs as a function of compatibility[15]. The percentage of incorrect or missing go-responses (PI) was determined accordingly. However, because of the rather strict error-based exclusion criterion (i.e., 10% across conditions, see above), the errors in the present set of experiments are less conclusive than in the dual task experiments reported above (see Chapter 4). Therefore, only overall (omnibus) PI analyses were run to ensure that the error pattern did not contradict RTs. Data were aggregated on compatibility conditions instead of stimulus position x response location primarily because of the low number of trials per condition. That is, aggregating on compatibility conditions seems to provide more reliable estimates. Moreover, this procedure seems more comparable to the aggregation procedure used in the dual task experiments presented in the previous chapter.

Again, in all analyses presented in the result section of Experiment 4 (and Experiment 5), reported p – values for effects involving factors with more than two within-'subjects' conditions are based on Greenhouse-Geisser corrected degrees of freedom. This also applies to the item analyses.

In addition to the subject analysis, an item analysis was carried out to test whether the same pattern of results emerges when items (instead of participants) are treated as a random factor. This type of analysis is akin to the item-analyses regularly found in psycholinguistic studies (e.g., Clark, 1973). Like in psycholinguistic studies, an item analysis seems warranted in the present experiment because (a) for a single subject, different compatibility conditions were realized by different items, and (b) a rather restricted number of items (i.e., letter pairs) realized each condition. Therefore, the interpretation of the results from the subject analysis rests on the assumption that the same results are obtained regardless of the specific item instances in a particular condition. If this assumption is correct, then similar results as in the subject analysis should be observed when items are treated as "subjects". For the item analy-

[15] The pattern of results for trimmed medians closely resembled the numerical pattern of medians based on unscreened RTs. Similarly, analyses on trimmed means led to comparable results as the analyses on trimmed medians reported here.

sis median RTs[16] for correct go-responses, and PIs for incorrect go-responses, were calculated as a function of letter pair (i.e., letters occurring together in a given S-R instruction) and compatibility, averaging across participants. Four items (letter pairs) had to be excluded from the item analysis because they served as no-go trials for the same compatibility condition on more than one list, and therefore did not provide data for all compatibility conditions. After excluding these four items, data from forty-four items, based on a maximum of sixteen observations (participants) per compatibility condition, were entered into the analysis. The motivation for treating letter pairs instead of single letters of a pair as items was similar to that of aggregating on compatibility instead of stimulus position x response location in the subject analysis, namely to maximize the number of observations per item and condition.

Main Results

The group means of the individual median RTs and the PIs for the final sample of subjects ($N = 24$) are presented in Table 11. As is evident from Table 11, there was a 19 ms Simon effect in RTs that was due to interference on incompatible trials.

Table 11. Mean Median Reaction Times (RT in ms) and percent invalid (PI) as a Function of Compatibility between Go-Signal Position and Response Location in Experiment 4.

	S-R compatibility			
	Compatible	Neutral	Incompatible	Δ
RT	332	332	351	19 ms
PI	1.9	3.0	4.3	2.4%

Note: The Column labeled Δ indicates the size of the Simon effect (incompatible minus compatible).

This observation received support from the omnibus RT ANOVA with S-R compatibility (compatible, neutral, incompatible) as a within-subjects factor, $F(2,46) = 10.02$, $p < .001$, $MSe = 270.01$. Planned comparisons testing all three 'components' separately, that is, the overall Simon effect (i.e., compatible vs. incompatible), its interference component (i.e., neutral vs. incompatible), and the facilitation component (i.e., neutral vs. compatible), showed that the Simon effect, $F(1,23) = 22.24$, $p < .001$, $MSe = 370.12$, and the interference component, $F(1,23) = 13.27$, $p < .01$, $MSe = 602.29$, were significant, whereas the facilitation component was not, $F(1,23) < 1$, $MSe = 647.65$.

The compatibility main effect in the overall PI ANOVA just missed significance, $F(1,46) = 2.8$, $p < .08$, $MSe = 12.19$. However, PIs showed a similar pattern as RT, as indi-

[16] Data aggregation for the item analysis was based on untrimmed raw RTs.

cated by an overall 2.4% Simon effect and the MANOVA that yielded a significant compatibility effect, $F(4,20) = 8.4$, $p < .001$.

Additional Analyses

As for the dual task experiments (cf. Chapter 4), a distribution analysis was carried out to explore the temporal dynamics of the Simon effect. However, unlike in the dual task experiments reported above and in the study by Hommel (1996c, Exp. 1), rank-ordered RTs were segregated into only two instead of five bins for each participant and condition. The reason to go with median splits (i.e., fast vs. slow RTs for each participant and compatibility condition) instead of a quintile analysis was that, for the former, the maximum number of observations per bin and condition was already reduced to 14, whereas the latter would be based on only 5-6 observations each, likely rendering the RT estimates highly unreliable. Figure 10 shows the group means of the median RTs for each participant, RT bin and compatibility condition.

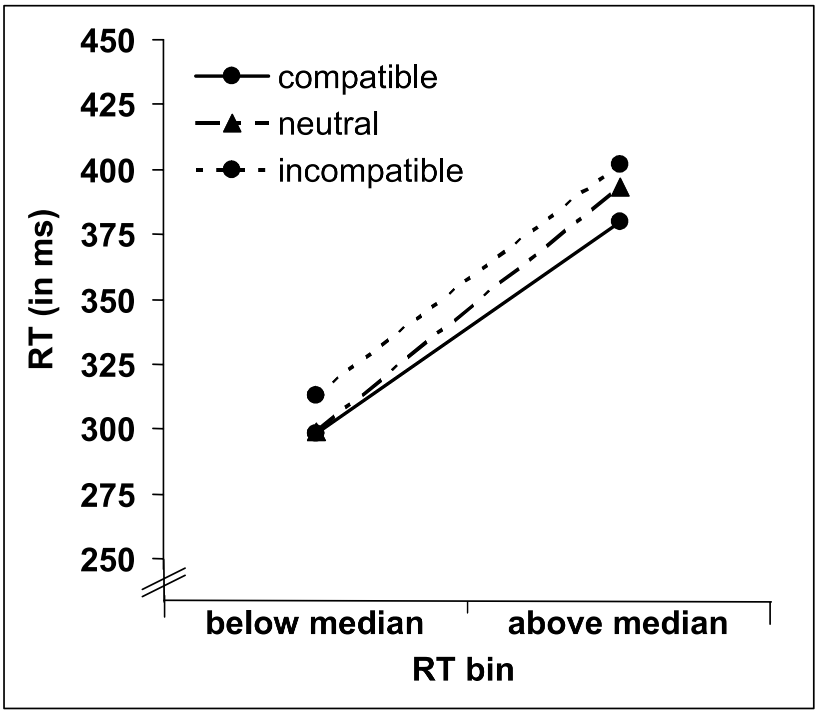

Figure 10. Mean median RTs for slow and fast responses as a function of compatibility between Go-Signal Position and Response Location in Experiment 4.

Figure 10 indicates that (a) the Simon effect was present for both fast and slow responses, but that (b) the effect (i.e., incompatible minus compatible) tended to be slightly larger for slow than for fast RTs, and (c) that the Simon effect for slow responses tended to be more symmetric with respect to interference (9 ms) and facilitation (13) ms than for fast responses (14 ms vs. 1 ms, respectively). However, the 2 (RT-bin) x 3 (compatibility) within-subjects ANOVA only yielded significant main effects of RT-bin, $F(1,23) = 260.34$, $p < .001$, $MSe = 1{,}089.47$, and compatibility, $F(2,46) = 5.82$, $p < .01$, $MSe = 720.28$, whereas the interaction of RT-bin and compatibility was not significant, $F(2,46) = 1.19$, $p > .3$, $MSe = 400.89$.

Finally, I also carried out an item-analysis to assess whether the same pattern as in the main analysis emerges when items instead of subjects are treated as a random factor. To this end, I determined the median RTs and PIs for each of the forty-four letter pairs for which data points existed for all compatibility conditions (cf. Section Data Analysis, above) as a function of compatibility (see Table 12 for means of medians across items). The within-'subjects' ANOVA of item RTs showed that the compatibility conditions significantly differed from each other, $F(2,86) = 6.85$, $p < .01$, $MSe = 1{,}257.67$, thus corroborating the results obtained in the subject analysis. Interestingly, in the item analysis, the Simon effect seemed to be more symmetric regarding the neutral condition than in the subject analysis. However, planned comparisons only revealed a significant Simon effect, $F(1,43) = 18.44$, $p < .001$, $MSe = 1{,}868.07$, whereas the interference and the facilitation component missed significance ($F(1,43) = 2.62$, $p > .11$, $MSe = 3{,}117.23$; and $F(1,43) = 3.54$, $p > .06$, $MSe = 2{,}560.74$, for interference and facilitation, respectively).

Importantly, the distribution of the Simon effect across items was clearly unimodal, and order correspondence (e.g., A: left key, N: right key as opposed to R: left key, C: right key) did not interact with compatibility, indicating that the Simon effect in the present experiment is not restricted to the subset of items with order-corresponding S-R assignments.

Table 12. Mean Median Reaction Times (RT) and percent invalid (PI) as a Function of Compatibility between Go-Signal Position and Response Location in the Item Analysis of Experiment 4.

	S-R compatibility			
	Compatible	Neutral	Incompatible	Δ
RT	328	342	356	28 ms
PI	1.9	3.3	4.2	2.3%

Note: The Column labeled Δ indicates the size of the Simon effect (incompatible minus compatible).

The pattern of PI across compatibility conditions closely resembled that in the subject analysis, and again missed significance, $F(2,86) = 2.42$, $p > .10$, $MSe = 24.35$. However, as in the subject analysis, the MANOVA on RT and PI showed a highly significant compatibility effect, $F(4,40) = 5.94$, $p < .001$, indicating that PI did not contradict RT in the item analysis.

5.1.3 Discussion

Experiment 4 extends the findings obtained by Hommel (1995, 1996c) as well as Shiu and Kornblum (1999) by demonstrating that a Simon effect of normal size (cf. Proctor & Lu, 1999, p. 67, who state that Simon effects observed with visual stimulus material are typically about 25 ms or less) occurs in go/no-go Simon tasks, even when the imperative stimuli are arbitrarily mapped to left and right responses and when S-R mappings change from trial to trial. This result implies that (a) the go-/no-go Simon effect does not depend on S-S congruency between a spatial precue and go-signal position (i.e., between go-position and direction of arrows in the Hommel experiments, or between go-position and location words in the Shiu & Kornblum, 1999, study), and that (b) the Simon effect in this task is not restricted to highly overlearned S-R mappings, but is also obtained when new stimuli need to be translated to their assigned spatially instructed responses on each trial. The results from the item analysis support the main results, indicating that the Simon effect observed in this Experiment is relatively stable despite the low number of items used in the experiment.

Numerically, the Simon effect tended to be slightly (7 ms) larger for slow than for fast responses, paralleling results obtained by Hommel (1996c, Exp. 1). Hommel proposed that faster responses exhibit a smaller Simon effect because at the time fast responses are emitted, the irrelevant location code has not yet been formed. In contrast, spatial codes are formed ‘in time’ to affect slow responses. However, unlike in the Hommel (1996c) experiment, the interaction between RT-bin and compatibility was not significant in the present study. Possible explanations are lack of power, and the relatively coarse segmentation of responses (i.e., 2 vs. 5 bins). The latter might have led to median RTs that differ less between bins than the average RTs in the extreme bins of Hommel’s quintile analyses. Moreover, RTs in my experiment were somewhat slower on average than in the Hommel experiment (this applies to medians as well as means), possibly resulting in more overlap in code activation in my experiment.

At least in the subject analysis, the compatibility effect in RTs was entirely due to interference (i.e., slower responses on incompatible than on neutral trials). No facilitation effect showed in the overall (subject) RTs. In contrast, most of the (relatively few) studies using a standard Simon task (in which the imperative stimuli appear at varying locations) that included a neutral condition found facilitation effects. Unfortunately, none of the experiments that used a go/no-go variant of the Simon task included a neutral condition. Thus, at present, it cannot be determined whether my results are akin to delayed stimulus-position type of tasks in general, or whether they are due to the specific neutral condition used in the present experiment. According to the former possibility, no facilitation effect might be observed in delayed-position experiments because facilitation only occurs with relatively slow responses (note that choice reactions are typically much slower, i.e., about 500 ms, than RTs in the present experiment), whereas interference is less affected by response speed or even more pronounced for fast responses. Some, albeit weak, evidence in support of this explanation comes from the RT-bin analysis that showed a (nonsignificant) facilitation effect for slow, but not for fast responses. This possibility is theoretically interesting because it may imply that response preparation, but not response initiation of fully prepared responses benefits from irrelevant corresponding location information.

However, there also was a tendency for a facilitation effect in the overall item RTs that cannot be explained in terms of differential response speed because overall RTs in the subject and the item analysis were almost identical. Hence, an alternative explanation of the present results is that there might have been much variability in how subjects treated the neutral condition, leading to inconsistent facilitation effects in the subject analysis when aggregating across items for individual subjects, but not necessarily so when averaging across subjects (i.e., in the item analysis). Clearly, further research is required to resolve this issue. One way of addressing this question would be to vary the SOA between imperative stimuli and go/no-go signals, possibly comparing different neutral conditions. If only unprepared responses benefit from corresponding trials, then larger facilitation effects should be observed at short than at long SOAs, regardless of how the neutral condition is realized.

A point related to the question of whether facilitation effects only occur for slow, unprepared responses concerns the locus of the Simon effect obtained with delayed location presentation. On the one hand, (a) the high proportion of go-trials (87.5%), (b) the fact that location information was presented 1700 ms after the onset of the imperative stimuli, and (c) the finding of comparatively short overall RTs, it seems safe to conclude that (imperative) stimulus identification and S-R translation had been completed on most trials when the go/no-go signals appeared, thus rendering a translation explanation of the Experiment 4 results unlikely. Instead, the Simon effect observed in this task strongly favors a direct response activation account. On the other hand however, one may ask whether the source of the Simon effect differs across tasks requiring unprepared vs. prepared responses. According to one view (e.g., Hommel, 1997), response selection is accomplished by activating the corresponding codes, and stimulus processing and response processing overlap in time. This view implies that

> […] as long as the response is not carried out, any response-congruent or conflicting stimulus information may facilitate or hamper responding. Consequently, response uncertainty should not play a major role […]. (Hommel, 1997, p. 298)

Accordingly, this view does not make a principled distinction between the locus of the Simon effect observed in situations where the response has to be prepared (i.e., when location information appears during S-R translation) and situations in which location information affects the initiation of a prepared response. However, one may disagree with this view if one adheres to more traditional, stage-like models. More specifically, it could be argued that the go/no-go Simon effect differs from the Simon effect observed when the position of the relevant stimulus varies in that the former reflects interference at initiating a prepared response, whereas the latter measures online interference during response selection (and, perhaps, response initiation; see, e.g., Shiu & Kornblum, 1999, for a discussion). I will return to this issue in the General Discussion section (Chapter 5.3).

In sum, Experiment 4 established a Simon effect for a 1-trial Simon task involving spatial response instructions while avoiding the confounding of instruction and practice effects typically present in other experiments on response coding. However, this experiment does not yet permit any conclusions about whether response instructions affect the Simon effect because spatial response instructions do not allow to discriminate between the alternative coding hypotheses (see Chapter 3). In order to address the question of whether response instructions affect response coding, response instructions in Experiment 5 were changed to non-spatial.

5.2 Experiment 5

Experiment 5 was largely identical to Experiment 4, except that responses were no longer instructed in terms of location. Rather, keys were instructed in terms of color, as in Experiments 2 and 3. That is, participants were instructed to press the green or the blue key, depending on letter identity. As in Experiment 4, a new letter pair was instructed on each trial, this time specifying letter to color (key) mappings (e.g., A: blue key, N: green key, on trial n; and R: blue key, C: green key, on trial n+1). Unlike the dual task Experiments 2 and 3, color-to-key assignment remained constant within a given block of trials, and was changed after a block had been completed. This modification was motivated by two considerations. First, keeping color-to-key assignment constant within a given block seemed to render the procedures in Experiments 4 and 5 more comparable. Second, it appeared to provide a fairer test of the different coding hypotheses because a constant color-to-(key) location assignment within a block can be assumed to facilitate recoding of responses in terms of location.

The predictions were similar to those of Experiment 3. If instruction determines response coding, that is, if participants include and weigh response codes in their response representations as instructed, then instructing responses in terms of color should deemphasize spatial response codes, and hence reduce the influence of irrelevant stimulus location. Therefore, the strong version of the direct coding hypothesis predicts a reduced Simon effect (reduced compared to the Simon effect with spatial response instructions).

In contrast, according to both the spatial and the weak version of the direct coding hypothesis, the spatial Simon effect under color instructions should not differ from that observed in Experiment 4. This is expected because both hypotheses assume that spatial coding is unaffected by instructions.

5.2.1 Method

Participants

Thirty-two students and non-students from Berlin (18 female, 14 male, mean age = 23 years) received € 7,- for participation (see Section Data Analysis, below, for a description of the final sample).

Stimulus Material and Counterbalancing

The same stimulus material and counterbalancing scheme as in Experiment 4 was used, with the following exceptions.

The S-R mapping instructions for each letter pair were changed such that letters were assigned to keys referred to by the color names BLAU and GRÜN, the German words for blue and green, respectively (e.g., R: blue key; C: green key). Color-to-key assignment was changed from block to block. On the practice block, as well as on Blocks 2 and 4, the green key was the one located on the left, and the blue key the one on the right. On the remaining blocks the assignment was reversed. This was true for all lists. Thus, letters that had been assigned left vs. right key responses in Experiment 4 were now assigned blue or green key responses, depending on block, keeping the physical letter-to-key (location) assignment constant across experiments. Because color-to-key assignment was changed between blocks, changes of letter-color assignments for a given letter on successive blocks (i.e., in new letter pairs) occurred when the location of the response remained the same. For 22 out of the 24 letters the color-assignment changed at least once across blocks.

The same trials as in Experiment 4 served as no-go trials, again “eliminating” two trials in each stimulus-position x response-location condition, leaving fourteen trials per condition (28 per compatibility condition) for the go-trial analyses. Because the same no-go trials were used as in Experiment 4, and because color-to-key assignment changed between blocks, the number of blue and green no-go trials was not equal on all lists. More specifically, whereas the same number of blue and green trials was (accidentally) excluded on Lists 1 and 3, on List 2 there were 8 green no-go trials (4 left and 4 right responses; 2 green trials on each block) and only 4 blue no-go trials (2 left and 2 right; 1 blue trial on each block).

Apparatus and Procedure

Apparatus and procedure were the same as in Experiment 4 with the following exceptions. In the introductory instructions, the position of the response keys was not mentioned. Rather, participants were told that their task would be to press either the blue or the green key, depending on the letter, and that they would be informed at the beginning of each block which key would be the blue viz. green key in that block.

Accordingly, each block started with a presentation of color patches of 3.5 x 3.5 cm size that simultaneously appeared at the left and right screen positions (i.e., the same lateral positions at which the go/no-go signals appeared). As noted above, on the practice block, as well as the second and fourth experimental block, the green patch appeared on the left, and the blue key appeared on the right, indicating that the left key had to be pressed when a green response was required and vice versa. On the first and third experimental blocks, the color-to-key assignment was reversed. When participants saw the color patches for a given block on the screen, the experimenter arranged color labels on the keys accordingly. These color labels consisted of 2.6 x 3.8 cm (height x width) paper color patches that lay behind the keys on the metal plates on which the keys were mounted. Extra color-labels on keys were provided to ensure that participants would not screw up a complete block of trials just because they had forgotten the color-to-key assignment.

Counterbalancing of lists[17], the sequence of trials, and the sequence of events within a trial were identical to Experiment 4, except that the mapping instructions (i.e., instructions assigning letters to responses) now instructed green vs. blue as opposed to left vs. right key responses. Substituting the spatial mapping instructions with color instructions resulted in presenting the green-key S-R mapping above the blue-key assignment on half of the trials in each block, whereas the order was reversed for the other half, again considering ordinal correspondence of letter-to-key location assignment.

5.2.2 Results

Data Analysis

Again, several participants from the total sample were excluded according to the same pre-experimentally defined criteria as in Experiment 4. Six participants produced more than 10% errors or misses. One additional participant was excluded because his or her overall RT exceeded the group mean by more than 4 standard deviations. Finally, one further participant had to be excluded because of a program error. However, the pattern of results for the excluded subjects did not statistically deviate from the results for included participants (see main results below). That is, although there seemed to be a hint of a Simon effect in PIs (but not in RTs), this tendency did not reach significance when data from excluded participants of Experiments 4 and 5 were combined ($N = 15$). On average, false alarms on no-go trials occurred on about three of the twelve no-go trials (25.4%) and were not analyzed further.

Trimming of correct go-RTs and aggregation of RTs and PIs for the remaining twenty-four participants (13 female, 11 male; mean age = 23 years) was the same as in Experiment 4. Screened correct responses were again excluded. They occurred on 2.6% of the trials and were evenly distributed across compatibility conditions. The item analysis was based on the same 44 items (i.e., letter pairs) as in Experiment 4 and was carried out accordingly.

Main Results

For the data of the remaining $N = 24$ participants, median RTs and PIs were computed for each participant and compatibility condition (see Table 13 for group means).

---

[17] Due to experimenter error, two participants who should have received one list according to the counterbalancing scheme were mistakenly given another list.

Table 13. Mean Median Reaction Times (RT in ms) and percent invalid (PI) as a Function of Compatibility between Go-Signal Position and Response Location in Experiment 5.

	S-R compatibility			
	Compatible	Neutral	Incompatible	Δ
RT	332	331	337	5 ms
PI	1.6	1.0	2.5	0.9%

Note: The Column labeled Δ indicates the size of the Simon effect (incompatible minus compatible).

Table 13 shows that the effects in both, RT and PI were numerically very small and extremely reduced as compared to the 19 ms Simon effect under spatial response instructions (i.e., in Experiment 4) The overall RT level was comparable across experiments (333.3 ms in Experiment 5 vs. 338.3 ms in Experiment 4).

These observations were supported by the ANOVAs. The compatibility effect in RTs did not reach significance, $F(2,46) = 1.25$, $p > .29$, $MSe = 226.61$, nor did any of the planned pairwise comparisons ($F(1,23) = 1.14$, $p > .29$, $MSe = 534.30$, for the Simon effect; and $F(1,23) < 1$, $MSe = 436.15$, and $F(1,23) = 2.66$, $p > .11$, $MSe = 389.19$, for facilitation and interference, respectively).

Compatibility did not reach significance in the omnibus PI ANOVA either, $F(2,46) = 1.88$, $p > .16$, $MSe = 7.25$. Although the PI pattern across compatibility conditions seemed to generally follow the RT pattern, additional correlation analyses and a MANOVA were carried out to ensure that the null results were not masked by some kind of speed-accuracy tradeoff. Neither analysis revealed a hint for a tradeoff. The correlation between the Simon effects (i.e., the Δs) in RTs and PIs was nonsignificant, $r = -.08$, $p > .7$. Similarly, the doubly multivariate MANOVA that considered all compatibility conditions again supported the RT analysis by not revealing a significant compatibility effect either, $F(4,20) = 1.19$, $p > .34$.

Further evidence for a reduction of the Simon effect under non-spatial compared to spatial response instructions comes from a comparison of the Simon effects in Experiments 4 and 5. The 2 (experiment) x 2 (compatible vs. incompatible) ANOVA on RTs revealed a significant main effect of compatibility, $F(1,46) = 14.7$, $p < .01$, $MSe = 226.10$, that was qualified by an interaction between compatibility and experiment, $F(1,46) = 4.83$, $p < .05$, indicating that the Simon effect was significantly larger in Experiment 4 than in Experiment 5. Moreover, as already noted, overall RT level was comparable across experiments, as indicated by a nonsignificant main effect of experiment, $F(1,46) < 1$, $MSe = 5{,}651.33$.

Additional Analyses

As in Experiment 4, a median-split analysis was carried out to assess whether the effect differed as a function of response speed. Figure 11 shows the group means of the medians of each participant, compatibility condition, and RT-bin.

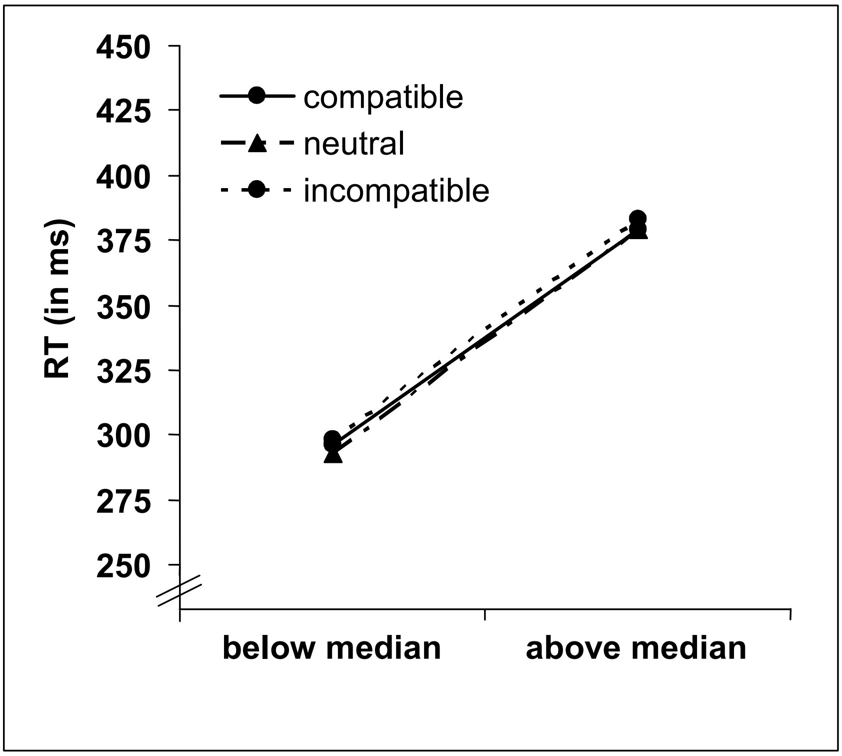

Figure 11. Mean median RTs for slow and fast responses as a function of compatibility between Go-Signal Position and Response Location in Experiment 5.

No compatibility effect was obtained at either RT-bin in Figure 11. The Simon effect was 2 ms for fast responses, and 4 ms for slow responses, indicating that the (null-) effect obtained in the main analysis did not vary with response speed. That is, no reversal of the effect at either RT-bin masked an effect at the other bin, as also indicated by the ANOVA that only yielded a main effect of RT-bin, $F(1,23) = 192.17$, $p < .001$, $MSe = 1{,}336.87$. Neither the compatibility main effect nor the interaction of compatibility and RT-bin reached significance, both F's < 1.

Finally, the item analysis again revealed that the pattern of results obtained for subjects generalizes to items, as indicated by the group means (of item medians) presented in Table 14 and the item ANOVAS.

Table 14. Mean Median Reaction Times (RT) and percent invalid (PI) as a Function of Compatibility between Go-Signal Position and Response Location in the Item Analysis of Experiment 5.

	S-R compatibility			
	Compatible	Neutral	Incompatible	Δ
RT	335	331	336	1 ms
PI	1.7	0.9	2.4	0.7%

Note: The Column labeled Δ indicates the size of the Simon effect (incompatible minus compatible).

The item RT ANOVA revealed that compatibility did not reach significance, $F(2,86) < 1$, $MSe = 454.74$. Nor did the Simon effect, or any of its components (all F's < 1) when analyzed separately in planned contrasts. As in Experiment 4, the distribution of the Simon effect in item RTs was unimodal, and compatibility did not significantly interact with order correspondence, again suggesting that order correspondence has a negligible effect in the present task.

The PI pattern in the item analysis closely resembled that of the subject analysis, and neither the PI ANOVA nor the MANOVA indicated any tradeoff between RT and PI (compatibility effects of $F(2,86) = 2.45$, $p > .1$, $MSe = 10.54$; and $F(4,40) = 1.55$, $p > .2$, for the PI ANOVA and the MANOVA, respectively).

5.2.3 Discussion

When responses were instructed in terms of color, the effect of corresponding viz. non-corresponding irrelevant position information on responding was small (5 ms) and not significant. Moreover, the Simon effect with non-spatial response instructions was significantly smaller than the 19 ms Simon effect observed with spatial response instructions (Experiment 4).

The results from the experimental comparison, as well as the fact that the statistical error term for the compatibility effect in the present experiment did not exceed that of Experiment 4 argues against a power explanation of the null-effect. Furthermore, the item analysis was consistent with the subject analysis. The item-RT distribution was again unimodal, and compatibility did not interact with alphabetic ordering, suggesting that the null-effect observed in Experiment 5 is “real” and cannot be attributed to a reversed effect in a subset of items (i.e., order-noncorresponding letter pairs).

Unlike the dual-task Experiment 3 reported above, color-to-key assignment remained constant throughout a complete block of trials, that is, recoding instructions and/or responses in terms of location was comparatively easy.

The Simon effect did not increase with increasing RT level, as indicated by the distribution analysis. This finding suggests that even less "prepared" responses, that is, responses for which S-R translation may not have been completed when the go-signal was presented, were hardly affected by irrelevant stimulus location under non-spatial response instructions.

Consequently, Experiment 5 corroborates the Experiment 3 results and provides converging evidence in favor of the strong direct coding hypothesis. It suggests that participants arbitrarily coded their responses in terms of color as instructed, weighing instructed codes more strongly than uninstructed spatial codes.

5.3 General Discussion Experiments 4-5

The experiments presented in Chapter 5 corroborate the dual task results reported in Chapter 4 by providing converging evidence for an impact of response instructions on <u>response</u> coding. More specifically, in Experiment 4 and 5, responses on a Simon-like task with delayed position presentation were either instructed spatially or in terms of color. Any effect of stimulus position and of response instructions on the Simon effect obtained with this task cannot easily be explained by translation accounts. Consistent with the intentional weighing hypothesis, which assumes that response codes can be weighed according to instructions and hence predicts a reduced Simon effect under non-spatial instructions, Experiments 4 and 5 showed that the 19 ms Simon effect observed under spatial response instructions was significantly reduced to 5 ms under color instructions.

The reduction of the Simon effect under color instructions contradicts dual route models such as the DO model proposed by Kornblum and colleagues (e.g., Kornblum et al., 1990; Kornblum et al., 1999; Zhang et al., 1999), which postulate that "when there is correspondence between the stimulus code and the response code, the latter is automatically activated, regardless of their relevance to the task" (Azuma et al., in press). Therefore, according to the DO model, which can be considered an example of the weak direct coding hypothesis, one would have expected a normal Simon effect under color instructions.

As a consequence, the results also contradict models that restrict activation via the direct route to the spatial dimension, and hence propose instruction-independent (pure) spatial response coding (e.g., De Jong et al., 1994).

Empirically, my results with non-spatial response instructions extend Hommel (1993a) who found a reversal of the Simon effect when response instructions emphasized different <u>spatial</u> aspects of the response array (i.e., either key location or contralaterally presented response effects). Unlike the Hommel (1993a) experiment, no salient visible action effects were presented in Experiment 5. Instead, color-plates were shown at the beginning of each block, and non-spatial response coding was evoked by verbally instructing the responses in terms of color before each trial. This finding indicates that instruction suffices to implement the intention to make color responses, that is, to implement arbitrary response codes in the action representation and persuade participants to use these codes in the control of responding.

The present findings also extend the results by Simon et al. (1976, Exp. 2; see Chapter 3.1.4) by providing a spatial response-instruction baseline for assessing the impact of non-spatial response instructions on the Simon effect. However, whereas the Simon effect in Experiment 5 was close to non-existent, Simon et al. observed a substantial (36 ms) effect even when color-to-key assignment changed from trial to trial. Several procedural differences may be responsible for this discrepancy. First, Simon et al. used an auditory variant of the Simon task that typically leads to much larger Simon effects (i.e., around 60 ms) than visual Simon tasks. Thus, as argued above, color instructions may have reduced the effect in Simon at al.'s varied label groups.

Moreover, auditory stimulus presentation, combined with the fact that lights attached to the keys served as color-labels in the Simon et al. experiment, enabled (or even required) subjects to fixate the response arrangement during stimulus processing. In contrast, participants could not simultaneously fixate the go/no-go signals (i.e., the screen) and the labels mounted on the response keys in my Experiment 5, presumably resulting in a higher likelihood of effector-position (or key location) coding in the former study. Finally, in the Simon et al. experiment, the relevant S-R mapping (i.e., pitch to color) was instructed only once at the beginning of the experiment, whereas new S-R mappings were instructed on each trial in Experiment 5. That is, in the latter but not in the former experiment, left and right keys were repeatedly referred to by color names, possibly priming the color dimension and reducing the likelihood of spatial re-coding.

Figure 12 illustrates how instructions in Experiments 4 and 5 might have affected response coding, and hence, the Simon effect, on incompatible trials.

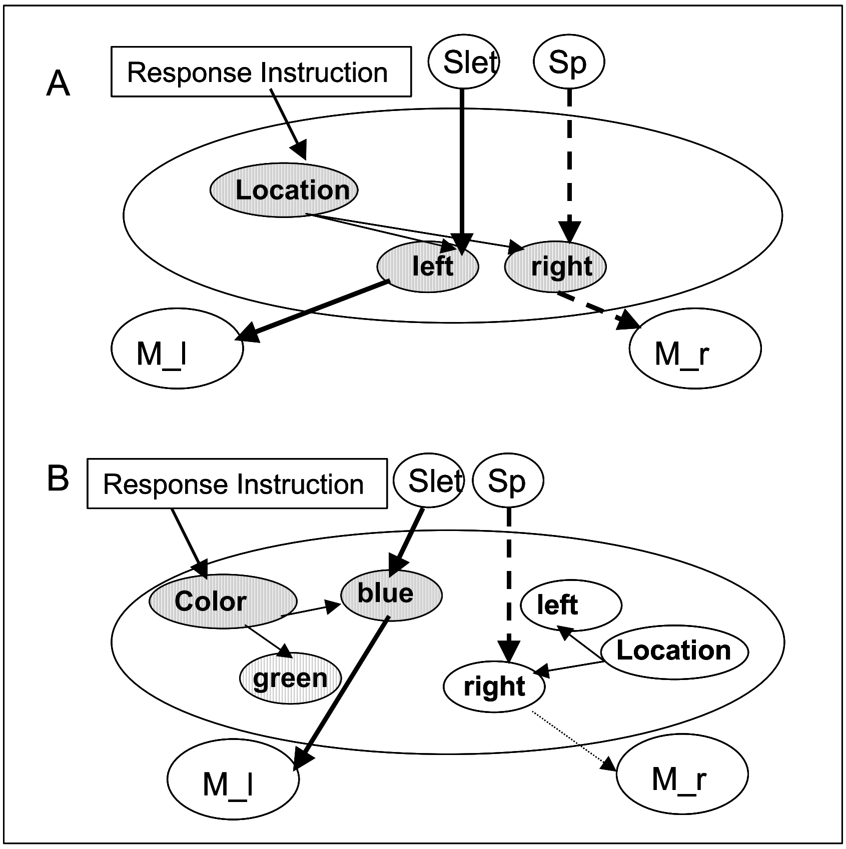

Figure 12. An illustration of the hypothetical activation flow on incompatible trials in Experiment 4 (Panel A) and Experiment 5 (Panel B). Go/no-go position (Sp) activates its corresponding location codes in either case. However, with non-spatial response instructions (Panel B) location codes are less strongly weighed, leading to less interference on incompatible trials. *Note*: Slet = (representation of) the imperative letter stimulus.

When responses are instructed in terms of location, as was the case in Experiment 4, spatial response codes are highly weighed and linked to/activated by the letter stimuli (Slet). Stimulus position (Sp) activates competing – highly weighed - location codes and thus leads to relatively strong response competition (Figure 12, Panel A). In contrast, when responses are instructed in terms of color, as was the case in Experiment 5, color codes are weighed more strongly than location codes that only play a minor role in response coding and selection. Consequently, even if stimulus position activates location codes (Figure 12, Panel B), this activation leads to less interference, and hence, reduced response competition.

Such an explanation of the reduced Simon effect under color instructions seems to bear some similarity with Magen and Cohen's (2002) notion of 'response-based input-selection'. In the experiments reported by Magen and Cohen, task-irrelevant flankers (and Stroop distractors) that were not part of the target set did or did not overlap with the outcome specification of <u>verbal</u> responses. These experiments showed distractor interference effects only when the distractors matched (or mismatched) the required verbal responses. Magen and Cohen explained these results by suggesting that the output specification primes specific diexplained these results by suggesting that the output specification primes specific dimensions, opening the door for irrelevant stimuli that overlap regarding the relevant response dimension. Similarly, in Experiments 4 and 5, a Simon effect was only observed when the irrelevant stimulus attribute overlapped with the instructed manual response dimension.

As already discussed (see Chapter 5.1.3), one peculiarity about the Simon effect obtained with spatial response instructions in Experiment 4 was that the RT effect was entirely due to interference, at least according to the subject analysis. Two possible reasons for this outcome have been discussed. According to one, this effect is an artifact produced by how subjects treated the neutral condition used here, whereas, according to the other explanation, an interference-dominant pattern of results may be systematically obtained with delayed-position type tasks. Although some researchers (e.g., Hommel, 1997) do not make a principled distinction between the Simon effects with vs. without response uncertainty, it is nevertheless possible that prepared responses (i.e., those for which S-R translation is completed at presentation of location information) are differently affected by irrelevant location information than yet-to-be-prepared responses. If so, one might argue that the observed instruction effect may be restricted to situations where fully programmed responses have to be initiated, but would not be as pronounced if irrelevant position information were presented during S-R translation.

While a definite answer to this question can only be provided by experiments that vary the SOA between the imperative stimuli and the go/no-go signals (see Section 5.1.3), the present results do not support this view. That is, if unprepared responses were more affected by stimulus position under non-spatial instructions, then there should have been a hint of a Simon effect for slow RTs in Experiment 5 because slow responses can be assumed to be less prepared on average. However, the distribution analysis of the Experiment 5 data did not indicate such a tendency – the Simon effect was 4 ms for slow as opposed to 2 ms for fast responses in Experiment 5. Therefore, it seems unlikely that the observed reduction of the effect was due to the specific task used here.

A final argument against an interpretation of the instruction effect reported in Chapter 5 in terms of response coding concerns recent observations that the use of the direct route might be – at least partially – under intentional control. For instance, Stürmer et al. (2002) demonstrated that the size of the Simon effect depends on (a) the proportion of compatible vs. incompatible trials, and (b) whether the preceding trial (i.e., trial n-1) was compatible or not.

Stürmer et al. found a reversed Simon effect when the overall percentage of incompatible trials was high. Moreover, they reported a regular Simon effect only for those trials following compatible trials, and this effect was independent of the overall proportion of compatible/incompatible trials. They interpreted the latter (sequence) effect as indicating suppression of the direct route after encountering conflict (i.e., upon incompatible trials), suggesting that the direct route may not be as unconditionally automatic as commonly assumed (but see Hommel, Proctor, & Vu, in press, for an alternative interpretation of such sequential effects).

Depending on how much intentional control (e.g., suppression) can be exerted on the direct route, one could argue that the reduction of the Simon effect in Experiment 5 was due to participants' deliberate decision not to use the direct route. However, this possibility seems questionable because Experiments 4 and 5 were identical in terms of the number of trials in each compatibility condition (all compatibility conditions were equally frequent), as well as regarding the sequence of trials. Thus, if participants had as much control over the direct route as implied by this view, the question arises as to why a Simon effect was observed in Experiment 4. More generally speaking, the question would be why an effect of irrelevant stimulus position (i.e., a Simon effect) usually shows with equiprobable compatibility conditions at all.

Nevertheless, in order to back up my conclusion that participants indeed coded their responses in terms of color, I am currently planning an experiment similar to Experiment 5, but with irrelevant color instead of location information, following the logic of Experiment 2. More specifically, responses will again be instructed in terms of color, and go/no-go signals determine whether a response is to be executed or not. However, unlike Experiment 5, go/no-go signals (vertical and horizontal bars) will be presented centrally and will randomly vary in terms of color instead of position. The (irrelevant) color of the go-signals either corresponds or does not correspond with the instructed key color, or is neutral with respect to the required response. If participants indeed arbitrarily code their responses as instructed, then irrelevant color information should automatically activate the highly weighed color codes, leading to interference on incompatible conditions. That is, a color-based Simon effect should be observed. Such an effect would also provide evidence against a strategic direct-route-suppression account of the reduced Simon effect in Experiment 5.

In sum, the Experiments presented in Chapter 5 extend existing findings and corroborate the dual-task results presented in Chapter 4 by showing that non-spatial response instructions can reduce the impact of irrelevant stimulus location on manual responding. In the present paradigm, irrelevant location information was presented considerably after the onset of the relevant (letter) stimulus, suggesting that the effect of instruction (i.e., the Simon effect in Experiment 4 and the lack thereof in Experiment 5) was not due to S-R translation in the conditional route. Rather, the findings seem to imply that instructions directly affected response coding. Thus, the present results provide additional evidence for the strong version of the direct coding hypothesis, which holds that codes referred to in the instructions are weighed more strongly and thus govern responding.

6 Conclusions

The general question motivating this thesis has been how task instructions are transformed into effective task sets that control instructed behavior. Although most researchers would probably agree that experimental instructions somehow determine how task sets are configured and therefore are important for the outcome of an experiment, relatively little is known about how exactly task instructions are compiled into task representations that are used to control behavior. The focus of this thesis has been on how the specific response labels used in the verbal instructions affect response coding in two-choice tasks involving spatially organized responses (i.e., left-right keypress responses). More specifically, the main question of this thesis has been whether the contents of response instructions directly determine how manual keypress responses are coded and accessed.

A promising way of addressing this question is to study the impact of response instructions on compatibility effects. This is so because compatibility effects are typically attributed to response priming that arises as a consequence of a match viz. mismatch between stimulus and response codes, or between response codes on two concurrently performed tasks. Consequently, investigating which match relations lead to compatibility effects under different instruction conditions allows conclusions about the cognitive codes that are used to control responding.

In Chapter 2, three theoretical positions have been discussed that differ regarding their assumptions on whether and how instructions affect coding of spatially organized responses, and hence with respect to their predictions concerning the nature and size of the compatibility effects under different response instructions.

According to the spatial coding hypothesis (e.g., De Jong et al., 1994), (response) instructions merely constrain how relevant stimulus attributes are mapped and translated to responses, without affecting response coding per se. Rather, this view assumes that responses are coded in terms of relative (i.e., left-right) key location whenever the spatial dimension allows discriminating between responses. Consequently, instruction-independent spatial compatibility effects of normal size and direction should be observed whenever response-overlapping spatial information is present or activated. Other than spatial compatibility effects should not occur as a function of overlap between task-irrelevant stimulus or concurrent response attributes that overlap with the instructed non-spatial response dimension. Rather, non-spatial compatibility effects should be restricted to the relevant S-R dimension, that is, they should be attributable to translation efficiency (i.e., some intermediate translation stage) in the conditional route.

This view has been contrasted with the direct coding hypothesis, which assumes that response labels directly influence response coding. According to this view, response labels activate their corresponding concepts that become included in the response representations and can be used to control responding. Because response-overlapping stimuli (or responses) are assumed to directly activate their corresponding responses, the direct coding hypothesis predicts compatibility effects resulting from overlap with the instructed response dimension, even when the instructed response dimension is non-spatial and the response-overlapping attribute is task-irrelevant.

Two versions of such a direct coding hypothesis have been distinguished with respect to spatially organized keypress responses. According to the weak version, as, for instance, represented by the DO model (e.g., Kornblum et al., 1990), top-down control of response coding is restricted. That is, instructed (non-spatial) codes cannot be weighed more strongly than uninstructed (spatial) codes. Consequently, this view makes similar predictions as the spatial coding hypothesis regarding spatial compatibility effects. More specifically, the weak direct coding hypothesis predicts that spatial compatibility effects are largely unaffected by (non-spatial) response instructions.

In contrast, according to the strong version of the direct coding hypothesis, the specific motor codes that are needed to perform the instructed response might primarily be accessible via the mental representation activated by the response label. According to this view that seems consistent with the intentional feature weighing hypothesis, (e.g., Hommel et al., 2001), it is primarily <u>intended</u> (instructed) action goals that are assigned and linked to attended features of stimuli. Hence, instructed (intended) stimuli and response features (that can be relatively abstract and non-spatial) are weighed more strongly than are irrelevant features, although the latter may still be part of the action representations. Accordingly, only the strong version of the direct coding hypothesis predicts that spatial compatibility effects are reduced under non-spatial response instructions.

The general conclusion drawn from the literature review (Chapter 3) on the impact of response instructions on a variety of compatibility effects (i.e., response coding) has been that results are highly inconclusive with respect to the different coding hypotheses, at least where non-spatial response instructions and response coding are concerned (see Chapters 3.1.5 and 3.3 for summaries).

In the empirical part of this thesis (Chapters 4 and 5), I therefore attempted to assess directly whether or not participants arbitrarily code their responses when so instructed, and whether non-spatial response coding can override spatial coding. The rationale underlying the experiments was to vary response instructions for manual (left and right) keypress responses to arbitrary stimulus attributes. This was done by instructing the response keys as either left vs. right keys (spatial instructions) or as blue vs. green keys (color instructions). Two experimental approaches were used to investigate whether and how instructions determine response coding. The first set of experiments (Experiments 1-3, Chapter 4) used a dual task procedure involving overlapping viz. non-overlapping responses on both tasks. In the second set of experiments (Experiments 4-5, Chapter 5), the dual task results were extended to a 1-trial-Simon type task with delayed position presentation.

In Experiments 1 and 4, spatially organized keypress responses were instructed spatially (i.e., by instructing the response keys as left and right), and overlapped with responses on a concurrently performed verbal task (i.e., “left” and “right” responses on the verbal task in Experiment 1) or with task-irrelevant position of go/no-go signals (Experiment 4). In both experiments substantial spatial compatibility effects were observed. Using the dual-task approach, Experiment 2 sought to generalize the spatial cross task compatibility effects to an arbitrary response dimension. To this end, manual responses as well as responses on the verbal task were instructed in terms of color. Substantial forward (i.e., verbal → manual) and backward (i.e., manual → verbal) color-based compatibility effects were observed. Finally, in Experiments 3 and 5 manual responses were again instructed in terms of color, but this time spatial coding was assessed. This was done by determining the compatibility effects resulting from ‘implicit’ overlap between non-spatially instructed manual keypress responses, on the one hand, with spatial concurrent responses (i.e., “left” and “right” verbal responses in Experiment 3) and with irrelevant stimulus position (Experiment 5) on the other hand. Both the spatial inter-task compatibility effects (Experiment 3) and the location-based Simon effect (Experiment 5) were found to be extremely reduced and statistically nonsignificant under non-spatial response instructions. These results have several important implications, which will be discussed in turn.

Implications for response coding. First, the backward and forward color compatibility effects between verbal and manual color responses under color instructions (Experiment 2) suggest that color instructions of manual responses, possibly assisted by repeated presentation of color patches, primed conceptual codes belonging to the color dimension that are or can be used in response selection. This finding extends demonstrations of inter-task consistency effects that used spatial response instructions and contradicts spatial coding accounts, which assume obligatory spatial coding regardless of response instructions (e.g., De Jong et al., 1994; Lu, 1997). Rather, it extends results on arbitrary code integration with practice by indicating that instructions may suffice to implement the intention to make color responses, thereby 'coloring' spatially organized keypress responses. Such a finding can be more easily explained by the two versions of the direct coding hypothesis according to which non-spatial (instructed) features can be used in the control of responding.

Experiments 3 and 5, on the other hand, suggest that color codes were not only part of the action representations, but that color coding can override spatial coding. More specifically, Experiments 3 and 5 suggest that color codes provide a viable alternative route to motor program activation (see Figure 13), and that codes can be weighed according to instructions. Accordingly, non-spatial response coding renders (irrelevant) spatial information less influential because spatial codes contribute less to responding. The results of Experiments 3 and 5 contradict coding accounts such as the DO model (e.g., Zhang et al., 1999) that can be considered instances of the weak direct coding hypothesis. Because these models assume comparable activation via the direct route for implicit and explicit (conceptual) overlap they predict spatial effects under color instructions. Rather, the results seem to support the intentional weighing hypothesis (e.g., Hommel et al., 2001) according to which intended (instructed) codes dominate how a response is represented and accessed. Therefore, response instructions seem to be at least in part responsible for how an otherwise identical (or very similar) task is performed, and whether (irrelevant) spatial information can be ignored or not. As a consequence, the present results also bear on issues of intentional control and automaticity, which will be discussed after some comments on my assumptions regarding the 'format' or nature of response codes (primarily) responsible for response selection according to my interpretation.

Some speculations in this regard seem in order to better relate the present interpretation to the existing literature, and to avoid confusion as to what I mean by 'conceptual' codes mediating response selection.

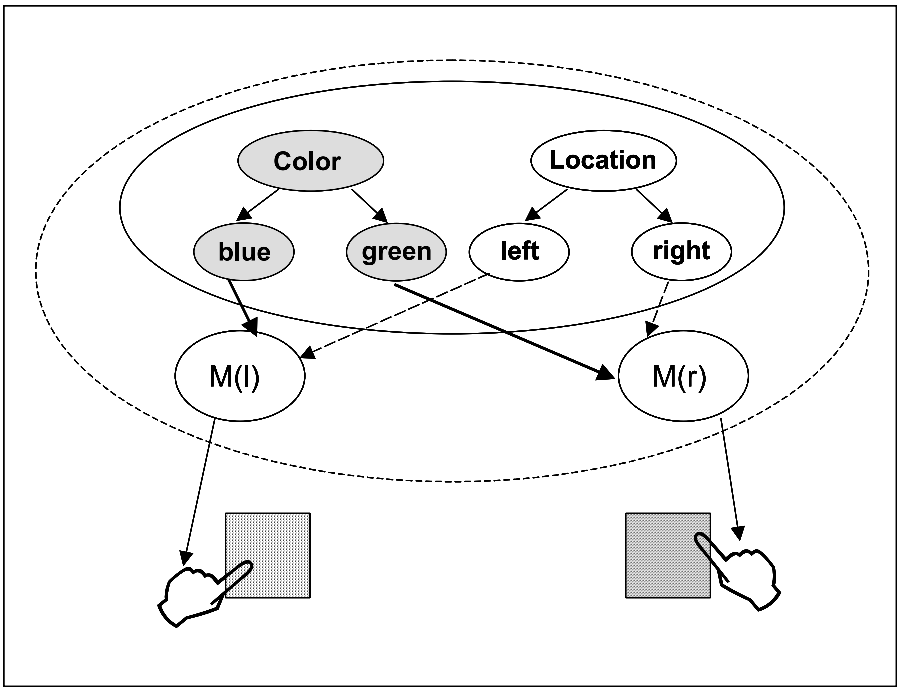

Figure 13. Sketch of the major theoretical implications regarding the impact of color instructions on response coding (adapted from Hommel, submitted). When keys are instructed in terms of color, color codes are integrated into the response representation. Instructions pre-activate the codes of a particular dimension (location or color), rendering spatial information less effective primes under color instructions (see text for details).

As noted above, my results and interpretation seem most consistent with the theory of event coding (TEC; Hommel et al., 2001). However, TEC rather explicitly assumes that stimuli and responses are coded in terms of distal <u>perceptually</u> based codes in a common representational medium. On the other hand, some researchers propose that color compatibility effects represent some sort of symbolic compatibility, implying that some type of symbolic codes that are often identified with the linguistic system contribute to the effect. Still others, including me, propose that such effects are largely conceptual, that is, meaning based. What I mean by ‘conceptual’ is that internal representations of the instructed categories become an integral part of the task sets and are used to control instructed responding. More specifically, I believe that instruction understanding involves both, extraction of propositional representations and construction of quasi-analogous situation models (e.g., Johnson-Laird, 1983). This implies that, in my view, category representations or meanings contained in the task sets are not only intensionally defined (i.e., with reference to other categories; e.g., *left* as meaning *¬right*), but also extensionally. That is, in terms of their referents in the real or represented world (cf. Johnson-Laird, Chaffin, & Herrmann, 1984; also see Barsalou, 1999). Viewed this way, conceptual coding can be considered at least partially perceptual or quasi-perceptual.

Inhowfar such a notion of conceptual coding relates to ‘symbolic’ or ‘verbal’ coding views is less clear. This is so because the latter terms seem to be very loosely defined and to be used in apparently different meanings within the compatibility literature. First, it appears as if the two terms are often used interchangeably. That is, ‘symbolic’ is equated with ‘verbal’. Second, ‘verbal’ coding is not consistently defined. For example, translation models of the Glaser and Glaser type (1989; see Chapter 3.1.4) seem to restrict the terms ‘verbal labels’ or ‘verbal system’ to purely lexical representations (i.e., concept <u>names</u>) that refer to semantic representations but do not represent semantics. The other extreme (i.e., ‘linguistic’ codes referring to purely semantic representations; e.g., Mattes et al., 2002) or some mixture of both (i.e., verbal codes containing names <u>and</u> some elementary semantics; e.g., Umiltà, 1991) has also been proposed. At present, I do not see a convincing theoretical basis or empirical support for the view that verbal (in the sense of lexical) codes substantially contribute to manual color responses once a task set is implemented, at least when keys are not labeled in terms of color <u>words</u> (see Chapter 3.1.4, for labeling effects in the manual Stroop task). Therefore, in my view, the distinction between symbolic and spatial compatibility lacks motivation. Rather, both types of compatibility should be considered conceptual (cf. Alluisi & Warm, 1990).

This is not to say that retrieval of concept names (i.e., inner speech) might not be helpful in concept activation during implementation or reconfiguration of S-R mappings (e.g., Emerson & Miyake, 2003; Goschke, 2000). However, I believe that verbal labeling processes mainly help to activate concepts, and thus may support implementation (and perhaps, consolidation) of S-R rules, but become less relevant once task sets have been implemented.

<u>Implications for intentional control and automaticity</u>. At a general level, the present results also bear on issues of intentional control and automaticity. On the one hand, they speak to the functional basis of what Luria (1961) called the ‘directive function of speech’, that is, how instructions come to control behavior. Luria demonstrated that the ability to recall instructions does not necessarily imply the ability to follow instructions. For instance, he observed that young children and patients with frontal lobe lesions, while being perfectly capable of understanding and recalling instructions, nevertheless have problems to consistently behave as instructed. That is, they show deficits in ‘controlled’ behavior that bear some similarity with what has become known as ‘goal neglect’ (e.g., Duncan, Emslie, Williams, Johnson, & Freer, 1996). Thus, it seems as if instruction following requires the ability to translate instructions into internal models that can be used to control behavior. The present results sug-

gest that instructions do not merely set up general constraints (e.g., by specifying the task-relevant stimulus category; see Chapter 2.1), but that the details or the specific contents of instructions (i.e., response instructions in the present study; but see, for example, Kunde, Kiesel, & Hoffmann, 2003, Exp. 3, for related findings concerning stimulus instructions) at least partially determine how internal models of the tasks are set up. More specifically,

> the resulting task set is likely to reflect the way the task is understood and interpreted by the perceiver/actor and, hence, determines how stimuli are coded (e.g., which stimulus features are attended and linked to response features) [and] how responses are coded (e.g., which response features are attended and linked to response features) […]. (Hommel, 2000, p. 266)

Regarding general models of action control, such as the Logan and Gordon (2001) model, this implies that response labels used in the instructions affect how parameter or parameter values are compiled from verbal instructions, and hence, how a task is performed. This should be considered by extensions of models like Logan and Gordon's, which still need to specify how verbal instructions are transformed into parameters, and which factors determine how this is done.

While the present results indicate that instructional details such as arbitrary instructed categories can be used in responding (that is, determine parameter values or pathways that cannot be assumed to be in the default response repertoire), at least when the relevant response categories are consistently primed throughout the experiment, future research needs to address under which conditions this conclusion does not hold. Such research will need to consider findings indicating that, in some situations, instructions are not or not consistently followed.

For example, the findings from the response-effect compatibility literature (cf. Chapter 3.1.3) indicate <u>non-instructed</u> response coding after practice by demonstrating the use of irrelevant (often arbitrary) codes that may have been primed through practice and/or which may have proven useful for the task at hand. Similarly, Kunde et al. (2003; Experiments 2 and 4) demonstrated that the internal model of the target (stimulus) set can be fine-tuned after (relatively little) practice. Kunde et al. found that masked priming was restricted to those stimuli from the instructed target categories that were actually experienced as targets whenever targets could easily be distinguished from non-targets.

Therefore, it is conceivable that non-spatially instructed responses might become (spatially) re-coded after practice under less optimal conditions than in my experiments (e.g., without repeated priming of the instructed response dimension).

In a similar vein, other findings suggest that the details of instructions are ignored or re-interpreted on some occasions. For instance, Prinz, Tweer, and Feige (1974; cited in Eimer, Nattkemper, Schröger, & Prinz, 1996) found that participants who had to detect certain targets (e.g., the letters 'A' and 'C') in a visual search task were slowed on, or even reported, pseudo-targets (i.e., letters that had not been defined by instructions and that are introduced after relatively little practice; e.g., the letter 'B'). This result indicates that participants performed the task by looking for items that deviate from their internal models of non-targets, rather than by matching the input to instruction-defined representations of the targets. Similarly, one possible explanation of the often observed interaction between compatibility effects in tasks with simultaneous S-R overlap on two dimensions (e.g., the two-dimensional spatial mapping task and the H&M task; see Chapters 3.1.1 and 3.1.4, respectively) is that subjects re-interpret instructions and perform these tasks by applying 'same' and 'different' rules (i.e., logical re-coding) to both task-relevant and irrelevant stimulus attributes.

In order to gain a more comprehensive understanding of how instructions are used to control behavior, and how instructed S-R rules are implemented within the cognitive system, future research will also need to generalize the present findings to more complex instructions and stimulus- and response arrangements. For instance, it should address whether other instructional factors than category labels, such as the specific stimulus and response examples given during instruction, the syntax of the instructions, and/or the order of mentioning also affect the contents of the resulting task set. That such factors might contribute to instruction understanding and task set configuration is suggested by findings from the text comprehension and problem-solving/reasoning literature (e.g., Johnson-Laird, Byrne, & Schaeken, 1992) on the one hand, and the learning literature (especially category learning and categorization; e.g., Nosofsky, Clark, & Shin, 1989), on the other hand.

In addition to providing insights into the functional basis of instructional and intentional control of behavior, the present work adds to and extends findings and reasoning on automaticity of S-R translation and/or response activation. More specifically, the present findings seem to fit in nicely with a 'prepared reflex' view of automaticity (see Hommel, 2000, for a comprehensive discussion), which holds that (a) once implemented, even arbitrary S-R rules are applied in an automatic (stimulus-triggered) fashion, but that (b) automatic response activation depends on how the task set is set up.

That is, the forward and backward compatibility effects in the dual task Experiments 1 and 2, as well as their lack of dependence on practice, add to the literature by showing that relatively little practice with arbitrary or even incompatible mapping leads to relatively strong automatic links that cannot be switched off when no longer needed (e.g., Hommel & Eglau, 2002; Proctor & Lu, 1999; Tagliabue et al., 2000). Second, Experiment 5 (and Experiment 3) adds to the evidence suggesting that the unconditional (direct) route is not as unconditionally automatic as sometimes assumed. Rather, instead of being primarily due to 'intrinsic' S-R strength (either hard-wired or highly overlearned; cf. Lu, 1997; Lu & Proctor, 2001), automatic response activation seems to depend on (a) how the intended responses are coded, (b) the readiness to respond with a particular key (Valle-Inclàn & Redondo, 1998), and (c) whether presented stimuli match the represented trigger conditions on the stimulus side (e.g., Kunde et al., 2003).

A look back and ahead. In sum, the present work addressed the questions whether and to what extent the response labels used in experimental task instructions determine how responses are coded, and hence how behavior is controlled. The results presented in this thesis suggest that research participants code and access responses as instructed even when response labels refer to arbitrary, non-spatial dimensions, presumably by activation and use of the category representations that correspond to the instructed labels. The findings imply a high degree of flexibility of coding that, in turn, determines which side effects (e.g., impact of irrelevant stimulus attributes) will be observed.

However, in order to gain a more comprehensive understanding of instruction following, or what Luria (1961) called the 'directive function of speech', future research needs to determine the constraining conditions of such labeling effects and to generalize such effects to more complex instructions as well as response arrangements (e.g., four-choice responses). For example, one question would be when and how simple S-R instructions are re-interpreted right away (e.g., in terms of same/different rules). Moreover, it will be interesting to see when and how learning modifies instructed responding. That is, under which conditions are 'instructed' task sets fine-tuned to task demands such that coding or weighing of codes is changed during practice? Of course, one would also need to address which processes (e.g., inner speech) afford implementation of S-R rules in the first place. Research along these lines will not only inform theorizing about how actions are intentionally controlled, but would also contribute to our understanding of how different types of automaticity depend on and relate to each other and instruction.

7 References

Allport, A. (1993). Attention and control: Have we been asking the wrong questions? In D. E. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV* (pp. 183-218). Cambridge, MA: MIT Press.

Alluisi, E. A., & Warm, J. S. (1990). Things that go together: A review of stimulus-response and related effects. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus-response compatibility: An integrated perspective* (pp. 3-30). Amsterdam: North Holland.

Azuma, R., Prinz, W., & Koch, I. (in press). Dual-task slowing and the effect of cross-task compatibility. *Quarterly Journal of Experimental Psychology (A)*.

Barber, P., & O'Leary, M. (1997). The relevance of salience: Towards an activational account of irrelevant stimulus-response compatibility effects. In B. Hommel & W. Prinz (Eds.), *Theoretical issues in stimulus-response compatibility* (pp. 135-172). Amsterdam: Elsevier.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-660.

Bashore, T. R. (1990). Stimulus-response compatibility viewed from a cognitive psychophysiological perspective. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus-response compatibility: An integrated perspective* (pp. 183-223). Amsterdam: North Holland.

Beckers, T., De Houwer, J., & Eelen, P. (2002). Automatic integration of non-perceptual action effect features: The case of the associative affective Simon effect. *Psychological Research/ Psychologische Forschung, 66*, 166-173.

Beringer, J. (2000). *Experimental Runtime System*. Frankfurt a. M.: Berisoft Corporation.

Botwinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624-652.

Brebner, J. (1979). The compatibility of spatial and non-spatial relationships. *Acta Psychologica, 43*, 23-32.

Broadbent, D. E., & Gregory, M. (1962). Donders' B- and C-reactions and S-R compatibility. *Journal of Experimental Psychology, 63*, 575-578.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335-359.

Cohen, J. D., Braver, T. S., & O'Reilly, R. (2000). A computational approach to prefrontal cortex, cognitive control, and schizophrenia: Recent developments and current challenges. In A. C. Roberts, T. W. Robbins & L. Weiskrantz (Eds.), *The prefrontal cortex: Executive and cognitive functions* (pp. 195-220). Oxford: Oxford University Press.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*, 332-361.

De Houwer, J., & Eelen, P. (1998). An affective variant of the Simon paradigm. *Cognition and Emotion, 12*, 45-61.

De Jong, R., Liang, C. C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual process model of effects of spatial stimulus response correspondence. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 731-750.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, P. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30*, 257-303.

Eimer, M., Nattkemper, D., Schröger, E., & Prinz, W. (1996). Involuntary Attention. In O. Neuman & A. F. Sanders (Eds.), *Handbook of Perception and Action* (Vol. 3, pp. 113-153). San Diego, CA: Academic Press.

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 229-240.

Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language, 48*, 148-168.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General, 118*, 13-42.

Goschke, T. (2000). Intentional reconfiguration and involuntary persistence in task set switching. In S. Monsell & J. Driver (Eds.), *Attention and Performance XVIII* (pp. 331-355). Cambridge, MA: MIT Press.

Guiard, Y. (1983). The lateral coding of rotations: A study of the Simon effect with wheel-rotation responses. *Journal of Motor Behavior, 15*, 331-342.

Hasbroucq, T., & Guiard, Y. (1991). Stimulus-response compatibility and the Simon effect: Toward a conceptual clarification. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 246-266.

Hedge, A., & Marsh, N. W. A. (1975). The effects of irrelevant spatial correspondences on two-choice response time. *Acta Psychologica, 39*, 427-439.

Heister, G., Schroeder-Heister, P., & Ehrenstein, W. H. (1990). Spatial coding and spatio-anatomical mapping: Evidence for a hierarchical model of spatial stimulus-response compatibility. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus-response compatibility: An integrated perspective* (pp. 117-143). Amsterdam: North Holland.

Hommel, B. (1993a). Inverting the Simon effect by intention: Determinants of direction and extent of effects of irrelevant spatial information. *Psychological Research/ Psychologische Forschung, 55*, 270-279.

Hommel, B. (1993b). The relationship between stimulus processing and response selection in the Simon task: Evidence for temporal overlap. *Psychological Research/ Psychologische Forschung, 55*, 280-290.

Hommel, B. (1995). Stimulus-response compatibility and the Simon effect: Toward an empirical clarification. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 764-775.

Hommel, B. (1996a). No prevalence of right-left over top-bottom spatial codes. *Perception and Psychophysics, 58*, 102-110.

Hommel, B. (1996b). The cognitive representation of action: Automatic integration of perceived action effects. *Psychological Research/ Psychologische Forschung, 59*, 176-186.

Hommel, B. (1996c). S-R compatibility effects without response uncertainty. *Quarterly Journal of Experimental Psychology, 49A*, 546-571.

Hommel, B. (1997). Toward an action-concept model of stimulus-response compatibility. In B. Hommel & W. Prinz (Eds.), *Theoretical issues in stimulus-response compatibility* (pp. 281-330). Amsterdam: Elsevier.

Hommel, B. (1998). Automatic stimulus-response translation in dual task performance. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1368-1384.

Hommel, B. (2000). The prepared reflex: Automaticity and control in stimulus-response translation. In S. Monsell & J. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes* (pp. 247-273). Cambridge, MA: MIT Press.

Hommel, B. (submitted for publication). Coloring an action: Intending to produce color events eliminates the Stroop effect. *Psychological Research/ Psychologische Forschung*.

Hommel, B., Alonso, D., & Fuentes, L. J. (in press). Acquisition and generalization of action effects. *Visual Cognition*.

Hommel, B., & Eglau, B. (2002). Control of stimulus-response translation in dual-task performance. *Psychological Research/ Psychologische Forschung, 66*, 260-273.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action. *Behavioral and Brain Sciences, 24*, 849-878.

Hommel, B., Proctor, R. W., & Vu, K.-P. L. (in press). A feature-integration account of sequential effects in the Simon task. *Psychological Research/ Psychologische Forschung*.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N., Byrne, R. M., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review, 99*, 418-439.

Johnson-Laird, P. N., Chaffin, R., & Herrmann, D. J. (1984). Only connections: A critique of semantic networks. *Psychological Bulletin, 2*, 292-315.

Keele, S. W. (1972). Attention demands of memory retrieval. *Journal of Experimental Psychology, 93*, 245-248.

Keele, S. W., Cohen, A., & Ivry, R. (1990). Motor programs: Concepts and issues. In M. Jeannerod (Ed.), *Attention and performance XIII: Motor representation and control* (pp. 77-110). Hillsdale, NJ: Lawrence Erlbaum Associates.

Koch, I., & Kunde, W. (in press). Verbal response-effect compatibility. *Memory and Cognition*.

Koch, I., & Prinz, W. (2002). Process interference and code overlap in dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 192-201.

Koch, I., & Prinz, W. (submitted for publication). Response preparation and code overlap in dual tasks.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus response compatibility - A model and taxonomy. *Psychological Review, 97*, 253-270.

Kornblum, S., & Lee, J.-W. (1995). Stimulus-response compatibility with relevant and irrelevant stimulus dimensions that do and do not overlap with the response. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 855-875.

Kornblum, S., Stevens, G. T., Requin, J., & Whipple, A. (1999). The effects of irrelevant stimuli: 1. The time course of stimulus-stimulus and stimulus-response consistency effects with Stroop-like stimuli, Simon-like tasks, and their factorial combinations. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 688-714.

Kunde, W. (2001). Response-effect compatibility in manual choice reaction tasks. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 387-394.

Kunde, W., Kiesel, A., & Hoffmann, J. (2003). Conscious control over the content of unconscious cognition. *Cognition, 88*, 223-242.

Lien, M.-C., & Proctor, R. W. (2000). Multiple spatial correspondence effects on dual task performance. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1260-1280.

Lien, M.-C., & Proctor, R. W. (2002). Stimulus-response compatibility and psychological refractory effects: Implications for response selection. *Psychonomic Bulletin & Review, 9*, 212-238.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review, 108*, 393-434.

Logan, G. D., & Schulkind, D. (2000). Parallel memory retrieval in dual-task situations: I. Semantic memory. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1072-1090.

Lu, C.-H. (1997). Correspondence effects for irrelevant information in choice-reaction tasks: Characterizing the stimulus-response relations and the processing dynamics. In B. Hommel & W. Prinz (Eds.), *Theoretical issues in stimulus-response compatibility* (pp. 85-117). Amsterdam: Elsevier.

Lu, C.-H., & Proctor, R. W. (1994). Processing of an irrelevant location dimension as a function of the relevant stimulus dimension. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 286-298.

Lu, C.-H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review, 2*, 174-207.

Lu, C.-H., & Proctor, R. W. (2001). Influence of irrelevant information on human performance: Effects of S-R association strength and relative timing. *Quarterly Journal of Experimental Psychology, 54A*, 95-136.

Luria, A. R. (1961). *The role of speech in the regulation of normal and abnormal behavior*. New York: Liveright Publishing Corporation.

MacLeod, C. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163-203.

Magen, H., & Cohen, A. (2002). Action-based and vision-based selection of input: Two sources of control. *Psychological Research/ Psychologische Forschung, 66*, 247-259.

Mattes, S., Leuthold, H., & Ulrich, R. (2002). Stimulus-response compatibility in intensity-force relations. *Quarterly Journal of Experimental Psychology, 55A*, 1175-1191.

McClain, L. (1983). Effect of response type and set size on Stroop color-word performance. *Perceptual and Motor Skills, 56*, 735-743.

Meiran, N. (2000). Reconfiguration of stimulus task sets and response task sets during task switching. In S. Monsell & J. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes* (pp. 377-399). Cambridge, MA: MIT Press.

Meiran, N. (2001). Event coding, executive control, and task-switching. *Behavioral and Brain Sciences, 24*, 893-894.

Milliken, B., Joordens, S., Merikle, P. M., & Seiffert, A. E. (1998). Selective attention: A reevaluation of the implications of negative priming. *Psychological Review, 105*, 203-229.

Monsell, S. (1996). Control of mental processes. In V. Bruce (Ed.), *Unsolved mysteries of the mind* (pp. 93-148). Hove, U.K.: Erlbaum.

Nicoletti, R., & Umiltà, C. (1984). Right-left prevalence in spatial compatibility. *Perception and Psychophysics, 35*, 333-343.

Nicoletti, R., & Umiltà, C. (1985). Responding with hand and foot: The right/left prevalence in spatial compatibility is still present. *Perception and Psychophysics, 38*, 211-216.

Nicoletti, R., Umiltà, C., Tressoldi, E. P., & Marzi, C. A. (1988). Why are left-right spatial codes easier to form than above-below ones? *Perception and Psychophysics, 43*, 287-292.

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 282-304.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 375-411). Cambridge: Cambridge University Press.

O'Reilly, R. C., & Soto, R. (2002). A model of the phonological loop: Generalization and binding. In T. G. Dietterich, S. Becker & Z. Ghahramani (Eds.), *Advances in Neural Information processing Systems (NIPS)* (Vol. 14). Cambridge, MA: MIT Press.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 16*, 220-224.

Phaf, R. H., Van der Heijden, A. H. C., & Hudson, P. T. W. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology, 22*, 273-341.

Prinz, W., Tweer, R., & Feige, R. (1974). Context control of search behavior: Evidence from a 'hurdling' technique. *Acta Psychologica, 38*, 72-80.

Pritchatt, D. (1968). An investigation into some of the underlying associative verbal processes of the Stroop colour effect. *Quarterly Journal of Experimental Psychology, 20*, 351-359.

Proctor, R. W., & Lu, C.-H. (1999). Processing irrelevant information: Practice and transfer effects in choice-reaction tasks. *Memory & Cognition, 27*, 63-77.

Proctor, R. W., & Pick, D. F. (2003). Display-control arrangement correspondence and logical recoding in the Hedge and Marsh reversal of the Simon effect. *Acta Psychologica, 112*, 259-278.

Proctor, R. W., Vu, K.-P. L., & Nicoletti, R. (2002). *A prevalence effect for two-dimensional stimulus-response sets when location is irrelevant?* Poster presented at the 43rd Annual Meeting of the Psychonomic Society, Kansas City, MO.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510-532.

Redding, G. M., & Gerjets, D. A. (1977). Stroop effect: Interference and facilitation with verbal and manual responses. *Perceptual and Motor Skills, 45*, 11-17.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*, 207-231.

Roswarski, T. E., & Proctor, R. W. (2000). Auditory stimulus-response compatibility: Is there a contribution of stimulus-hand correspondence? *Psychological Research/ Psychologische Forschung, 63*, 148-158.

Roswarski, T. E., & Proctor, R. W. (2003a). The role of instructions, practice, and stimulus-hand correspondence on the Simon effect. *Psychological Research/ Psychologische Forschung, 67*, 43-55.

Roswarski, T. E., & Proctor, R. W. (2003b). Intrahemispherical activation, visuomotor transmission, and the Simon effect. Comment on Wascher et al. (2001). *Journal of Experimental Psychology: Human Perception and Performance, 29*, 152-158.

Schuch, S., & Koch, I. (submitted for publication). The costs of changing the representation of action: Response repetition and response-response compatibility in dual tasks.

Sharma, D., & McKenna, F. P. (1998). Differential components of the manual and verbal Stroop tasks. *Memory & Cognition, 26*, 1033-1040.

Shiu, L.-P., & Kornblum, S. (1999). Stimulus-response compatibility effects in go-no-go tasks: A dimensional overlap account. *Perception and Psychophysics, 61*, 1613-1623.

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology, 81*, 174-176.

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus-response compatibility: An integrated perspective* (pp. 183-223). Amsterdam: North Holland.

Simon, J. R., Acosta, E., Mewaldt, S. P., & Speidel, C. R. (1976). The effect of an irrelevant directional cue on choice reaction time: Duration of the phenomenon and its relation to stages in processing. *Perception and Psychophysics, 19*, 16-22.

Simon, J. R., Hinrichs, J. V., & Craft, J. L. (1970). Auditory S-R compatibility: Reaction time as a function of ear-hand correspondence and ear-response-location correspondence. *Journal of Experimental Psychology, 86*, 97-102.

Simon, J. R., Sly, P. E., & Vilapakkam, S. (1981). Effect of compatibility of S-R mapping on reaction toward the stimulus source. *Acta Psychologica, 47*, 63-81.

Simon, J. R., & Sudalaimuthu, P. (1979). Effects of S-R mapping and response modality on performance in a Stroop task. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 176-187.

Smith, P., & Brebner, J. (1983). S-R compatibility: The relative effects of "relevant" spatial and non-spatial variables. *Australian Journal of Psychology, 35*, 1-10.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Stürmer, B., Leuthold, H., Soetens, H., Schröter, H., & Sommer, W. (2002). Control over location-based response activation in the Simon task: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 1345-1363.

Sugg, M. J., & McDonald, J. E. (1994). Time course of inhibition in color-response and word-response versions of the Stroop task. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 647-675.

Tagliabue, M., Zorzi, M., Umiltà, C., & Bassignani. (2000). The role of long-term-memory and short-term-memory links in the Simon effect. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 648-670.

Umiltà, C. (1991). Problems for the salient-features coding hypothesis: Comment on Weeks and Proctor. *Journal of Experimental Psychology: General, 120*, 83-86.

Umiltà, C., & Nicoletti, R. (1990). Spatial stimulus-response compatibility. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus-response compatibility: An integrated perspective* (pp. 89-116). Amsterdam: Elsevier.

Valle-Inclán, F. (1996). The Simon effect and its reversal studied with event-related potentials. *International Journal of Psychophysiology, 23*, 41-53.

Valle-Inclán, F., & Redondo, M. (1998). On the automaticity of ipsilateral response activation in the Simon effect. *Psychophysiology, 35*, 366-371.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory & Cognition, 13*, 304-319.

Vu, K.-P. L., & Proctor, R. W. (2001). Determinants of right-left and top-bottom prevalence for two-dimensional spatial compatibility. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 813-828.

Vu, K.-P. L., & Proctor, R. W. (2002). The prevalence effect in two-dimensional stimulus-response compatibility is a function of the relative salience of the dimensions. *Perception and Psychophysics, 64*, 815-828.

Vu, K.-P. L., Proctor, R. W., & Pick, D. F. (2000). Vertical versus horizontal spatial compatibility: Right-left prevalence with bimanual responses. *Psychological Research/ Psychologische Forschung, 64*, 25-40.

Wallace, R. J. (1971). S-R compatibility and the idea of a response code. *Journal of Experimental Psychology, 88*, 354-360.

Wang, D.-Y., Proctor, R. W., & Pick, D. F. (2002). *The Simon effect with wheel rotation response.* Poster presented at the 43rd Annual Meeting of the Psychonomic Society, Kansas City, MO.

Wascher, E., Schatz, U., Kuder, T., & Verleger, R. (2001). Validity and boundary conditions of automatic response activation in the Simon task. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 731-751.

Watter, S., & Logan, G. (2001). *Parallel response-selection in dual-task situations.* Poster presented at the 42nd Annual Meeting of the Psychonomic Society, Orlando, Florida.

Wenke, D., & Frensch, P. A. (2000). *From instruction to action.* Poster presented at the 41st Annual Meeting of the Psychonomic Society, New Orleans, Louisiana.

White, B. W. (1969). Interference in identifying attributes and attribute names. *Perception and Psychophysics, 6*, 166-168.

Worringham, C. J., & Kerr, G. K. (2000). Proprioception and stimulus-response compatibility. *Quarterly Journal of Experimental Psychology, 53A*, 69-83.

Zhang, H. (2000). The Simon effect and its reversal in three-choice Hedge and Marsh tasks: Evidence for irrelevant stimulus-response compatibility and stimulus congruity. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1018-1037.

Zhang, H., Zhang, J., & Kornblum, S. (1999). A parallel distributed processing model of stimulus-stimulus and stimulus-response compatibility. *Cognitive Psychology, 38*, 386-432.

Zhang, J., & Kornblum, S. (1997). Distributional Analysis and De Jong, Liang, and Lauber's (1994) Dual-Process model of the Simon effect. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1543-1551.

Ziessler, M., & Nattkemper, D. (2002). Effect anticipation in action planning. In W. Prinz & B. Hommel (Eds.), *Attention and Performance XIX* (pp. 645-672). Oxford: Oxford University Press.